Development and application of model selection methods for investigating brain function

Rosa, Maria

Awarding institution:
UCL University College London

Development and application of model selection methods for investigating brain function

Maria João Duarte Rosa

Wellcome Trust Centre for Neuroimaging

University College London

A thesis submitted for the degree of

Doctor of Philosophy

January, 2012

Primary supervisor: Dr. William D Penny

Secondary supervisor: Dr. James M Kilner

To my grandmothers.

Às minhas avós,
Sofia Duarte e Lucila Miguéns.

Acknowledgements

I owe sincere and earnest thankfulness to my supervisor Will Penny. Throughout my PhD, Will has been extremely supportive, patient and always available for discussion. I have learned immensely in our meetings and truly enjoyed working with Will. Will has been a constant source of encouragement, knowledge, and friendship with a brilliant sense of humour.

I would also like to thank my second supervisor, James Kilner for his enthusiasm, ideas, advice and crucial contributions to my PhD.

The work I have done would not have been possible without my co-authors. I thank Sven Bestmann, Lee Harrison and Jean Daunizeau for providing ideas, data and a lot of enthusiasm. I also thank Felix Blankenburg and Oliver Josephs for acquiring the data that I worked with during most of my PhD.

I would also like to show my gratitude to Karl Friston for critical contributions and inspiring discussions.

I sincerely thank everyone in the Methods Group, in particular Guillaume Flandin, Justin Chumbley, Klaas Stephan, Baojuan Li and Carlton Chu. I have learned immensely by attending the Methods II meetings and made lots of good friends.

I am truly thankful to Marta Garrido and Andre Marreiros. I am very lucky to have met Marta and Andre when I joined the lab. Since then, they have been my best friends and Portuguese family in London.

I also owe a lot of gratitude to Wako Yoshida and Ben Seymour for being my home during most of my PhD.

I would also like to thank all the fellows in the FIL, specially Zoltan Nagy and everyone on the second floor, for providing the best possible work environment. We have shared lots of good moments.

I would like to thank Peter Aston and Marcia Bennett for all the support with travel arrangements and other administrative work.

I would also like to show my gratitude to Eduardo Ducla-Soares, for the constant support and friendship since my undergrad years.

Finally, I would like to thank all my friends and family.

This work was funded by the Wellcome Trust and the Portuguese Foundation for Science and Technology.

Declaration

I, Maria João Duarte Rosa, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The goal of any scientific discipline is to learn about nature, usually through the process of evaluating competing hypotheses for explaining observations. Brain research is no exception. Investigating brain function usually entails comparing models, expressed as mathematical equations, of how the brain works. The aim of this thesis is to provide and evaluate new model comparison techniques that facilitate this research. In addition, it applies existing comparison methods to disambiguate between hypotheses of how neuronal activity relates to blood flow, a topic known as neurovascular coupling.

In neuroimaging, techniques such as functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) allow to routinely image the brain, whilst statistical frameworks, such as statistical parametric mapping (SPM), allow to identify regionally specific responses, or brain activations. In this thesis, SPM is first used to address the problem of neurovascular coupling, and compare different putative coupling functions, which relate fMRI signals to different features of the EEG power spectrum. These inferences are made using linear models and a model selection approach based on F-tests. Although valid, this approach is restricted to nested models. This thesis then focuses on the development of a Bayesian technique, to construct posterior model probability maps (PPMs) for group studies. PPMs are analogous to F-tests but not limited to nested hypotheses.

The work presented here then focuses again on neurovascular coupling, this time from a mechanistic perspective, not afforded by linear models. For this purpose, a detailed biophysical framework is used to explore the contribution of synaptic and spiking activity in the generation of hemodynamic signals in visual cortex, using simultaneous EEG-fMRI. This approach is a special case of brain connectivity models. Finally, using fMRI data, this thesis validates a recently proposed Bayesian approach for quickly comparing large numbers of connectivity models based on inverting a single model.

Contents

List of Figures

Abbreviations

AIC	Akaike Information Criterion
AIS	Annealed Importance Sampling
ANOVA	Analysis of Variance
AW	Wakefulness
BEM	Boundary Element Method
BF	Bayes Factor
BIC	Bayesian Information Criterion
BMA	Bayesian Model Averaging
BMS	Bayesian Model Selection
BOLD	Blood Oxygenation Level Dependent
CBF	Cerebral Blood Flow
CBV	Cerebral Blood Volume
$CMRO_2$	Cerebral Metabolic Rate of Oxygen Consumption
DCM	Dynamic Causal Modelling
dHb	deoxy-Hemoglobin
ECD	Equivalent Current Dipole
ECoG	Electrocorticography
EEG	Electroencephalography
EM	Expectation-Maximisation
EPI	Echo-planar Imaging
EPM	Exceedance Probability Maps
FDR	False Discovery Rate
FEM	Finite Element Method
FFX	Fixed Effects
fMRI	functional Magnetic Resonance Imaging

FWE	Family Wise Error
GBF	Group Bayes Factor
GFP	Global Field Power
GLM	General Linear Model
HRF	Hemodynamic Response Function
ICA	Independent Component Analysis
iEEG	Intracranial Electroencephalography
KL	Kullback-Leibler
LEV	Local Electro-Vascular
LFP	Local Field Potential
MAP	Maximum a Posteriori
MCMC	Markov Chain Monte Carlo
MDEFT	Modified Driven Equilibrium Fourier Transform
MEG	Magnetoencephalography
MF	Mean Frequency
MNI	Montreal Neurological Institute
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MUA	Multi-Unit Activity
NMM	Neural Mass Model
NO	Nitric Oxide
OHb	Oxy-Hemoglobin
PCA	Principal Component Analysis
PET	Positron Emission Tomography
ROI	Region of Interest
PPM	Posterior Probability Mapping
REM	Rapid Eye Movement
ReML	Restricted Maximum Likelihood
RF	Radio Frequency
RFX	Random Effects
RFT	Random Field Theory
RMS	Root Mean Squared
RMSF	Root Mean Squared Frequency
SEM	Structural Equation Modelling

SNR	Signal-to-Noise Ratio
SPC	Superior Parietal Cortex
SPM	Statistical Parametric Mapping
SSVER	Steady State Visual Evoked Responses
SUA	Single-Unit Activity
SVC	Small Volume Correction
SVC	Singular Value Decomposition
SVM	Support Vector Machine
TR	Echo Time
TO_2	Tissue Oxygenation Concentration
TR	Repetition Time
V1	Primary Visual Cortex
V5	Middle Temporal Visual Area
VB	Variational Bayes
VBM	Voxel-Based Morphometry
VL	Variational Laplace

Outline and aims of this thesis

The aim of this thesis was to develop and apply model selection methods for investigating brain function using human neuroimaging data. In particular, this thesis focuses on the relationship between neuronal activity and blood flow, also known as neurovascular coupling. Here, classical and Bayesian model selection methods are used to disambiguate between competing hypotheses for neurovascular coupling, using combined electroencephalographic (EEG) and functional magnetic resonance imaging (fMRI) data. In addition, this thesis presents a novel Bayesian technique to construct posterior model probability maps (PPMs) for making inferences about regionally specific effects using imaging data from a group of subjects. Finally, the work presented here explores the relation between different approximations to the model evidence in the context of deterministic dynamic causal models (DCMs), commonly used in brain connectivity analyses.

This thesis is organised as follows:

Chapter 1 - Introduction

This chapter introduces the brain imaging techniques and modelling frameworks used in this thesis, and situates them in the wider context of imaging neuroscience and brain research.

Chapter 2 - Methods

This chapter focuses on model selection methods used in neuroimaging, and with which the results in this thesis have been obtained.

Chapters 3 to 6 - Results

These chapters describe the original work developed in this thesis, including novel methodologies and experimental findings. In particular:

- **Chapter 3** explores the dependency of the BOLD signal on the temporal dynamics of the underlying neuronal activity. This chapter compares different frequency-dependent and -independent coupling (transfer) functions from neuronal activity to BOLD, using simultaneous EEG-fMRI from healthy subjects;
- **Chapter 4** describes the construction of PPMs for Bayesian model selection (BMS) at the group level. This method is illustrated using fMRI data from a group of subjects performing a target detection task;
- **Chapter 5** presents a modelling framework that can be used to non-invasively compare biophysically plausible neurovascular coupling mechanisms using EEG and fMRI data. This framework is used to investigate the role of synaptic versus spiking activity in the generation of BOLD signals in human visual cortex;
- **Chapter 6** presents and validates a recently proposed Bayesian approach for quickly comparing large numbers of deterministic DCMs, using fMRI data from an attention to visual motion study.

Chapter 7 - Discussion and conclusions

This chapter provides a general discussion and conclusions of the work presented in this thesis, and points out directions for future research.

Chapter 1

Introduction

The aim of this chapter is to introduce the measuring and modelling techniques used in this thesis, and situate them in the wider context of imaging neuroscience and brain research.

In brief, this chapter first reviews two of the primary techniques used for non-invasive imaging of the human brain: functional magnetic resonance imaging (fMRI) and electroencephalography (EEG). These techniques are non-invasive and, in general, widely available, allowing routine acquisition of human brain imaging data. Moreover, these two techniques have complementary advantages and limitations, and the recent years have seen a trend for integrating the two. EEG-fMRI integration is particularly useful for investigating the relationship between neuronal activity and blood flow, a topic known as neurovascular coupling. This chapter provides a brief review of some of the most important findings in this field.

In addition to the measuring techniques, this chapter introduces the modelling frameworks that are most commonly used to analyse such data. It first goes through basic statistical models used for making classical and Bayesian inferences about regionally specific responses, known as statistical parametric mapping (SPM) and posterior probability mapping (PPM), respectively. This chapter then focuses on how these basic models can be finessed, by incorporating biophysical constraints, to allow one to infer about hidden neurophysiological mechanisms. Finally, this chapter briefly reviews how these models can be used to infer how interactions among brain regions are mediated, known as effective connectivity.

1.1 Measuring brain function

1.1.1 Imaging techniques

This section aims to introduce two of the most commonly used techniques in imaging neuroscience: fMRI and EEG. This introduction focuses on the physiological origin of the signals, rather than on the hardware and data acquisition. The objective is to review the main advantages and limitations intrinsic to each measure, and motivate the integration of EEG and fMRI. An understanding of the physiological origin of the signals is important for the investigation of neurovascular coupling, as described in Chapters 3 and 5.

1.1.1.1 functional Magnetic Resonance Imaging (fMRI)

fMRI is a non-invasive and widely available technology that allows one to assess changes in functional brain activity recorded using a magnetic resonance (MR) scanner. The most commonly used fMRI contrast is the Blood Oxygenation Level Dependent (BOLD) signal, which is due to the hemodynamic and metabolic outcome of neuronal responses [Huettel et al., 2009].

It is well known that an increase in neuronal activity is accompanied by an increase in the cerebral metabolic rate of oxygen consumption ($CMRO_2$) and a much larger increase in local cerebral blood flow (CBF) [Frostig et al., 1990; Riera et al., 2008]. Due to this imbalance, local venule blood is more oxygenated following activation. It is also well known that oxygenated hemoglobin, oxyhemoglobin, is diamagnetic, while deoxygenated hemoglobin, deoxyhemoglobin, is paramagnetic [Pauling and Coryell, 1936]. This increase in oxygenated blood is therefore responsible for a reduction in the local distortions of the static MR field, which constitutes the BOLD signal [Kwong et al., 1992; Ogawa et al., 1992].

The BOLD signal is measured using rapid volumetric MR acquisition sequences [Jezzard et al., 2001] and in consequence has relatively good spatial resolution (Figure 1.1): voxels (3-D elements) in the image typically represent cubes of tissue of approximately 2 to 4 millimeters on each side. Higher resolution imaging is becoming routinely possible due to the advent of high field strength scanners (such as 7 Tesla) [Krebs et al., 2010; Yacoub et al., 2008]. In addition, arteriolar control of blood flow is known to be spatially well-matched to local increases in neuronal activity [Turner and Jones, 2003]. This means that the main application of fMRI is

mapping brain function, i.e. identifying functionally segregated regions in the brain responsible for sensorimotor and mental processes.

However, BOLD-fMRI is only an indirect measure, through metabolism, oxygenation and blood flow, of underlying neuronal activity. These slow mechanisms provide only temporally smoothed functions of this activity. Even though it is possible to distinguish the BOLD response to experimental stimuli presented several hundred milliseconds apart (in the same brain region), the time course of the response to a brief stimulus lasts at least 15 seconds [Huettel et al., 2009]. For these reasons, fMRI's temporal resolution is relatively poor compared to electrophysiological measures, such as EEG (described in the following section), albeit being better than other metabolic-based techniques such as Positron Emission Tomography (PET) [Poeppel, 1996] (Figure 1.1).

The main limitation of fMRI however, is the fact that the mechanisms of coupling between neuronal activity and blood flow (i.e. neurovascular coupling) are not fully understood [Logothetis and Pfeuffer, 2004]. Therefore no precise inferences about the underlying neuronal activity can be directly drawn from the fMRI signal. This issue is also referred to as the inverse fMRI problem. A large body of research has been devoted to understanding neurovascular coupling from different perspectives [Riera and Sumiyoshi, 2010]. For instance, this research aims to find out which aspects of neuronal activity (e.g. synaptic or spiking activity), and which cells (e.g. neurons or glial cells) and particular molecules (e.g. nitric oxide) are responsible for the observed changes in blood flow following neuronal activation. These mechanisms are discussed in more detail in a later section (see Neurovascular Coupling section, 1.1.2.2).

In summary, fMRI with BOLD contrast is an established method for making inferences about regionally specific activations in the brain. Its spatial localisation power is its main advantage relative to other techniques, but the lack of knowledge about the mechanics that couple underlying neuronal activity to blood flow hinders the interpretation of BOLD maps and how these maps relate to other measures of neuronal activity. This area is still a topic of much debate (see below).

1.1.1.2 Electroencephalography (EEG)

EEG, and magnetoencephalography (MEG), measure extra-cranial electrical potentials, and magnetic fields, respectively, which are known to be generated by brain current source activity [Hämäläinen et al., 1993; Niedermeyer and Lopes da Silva,

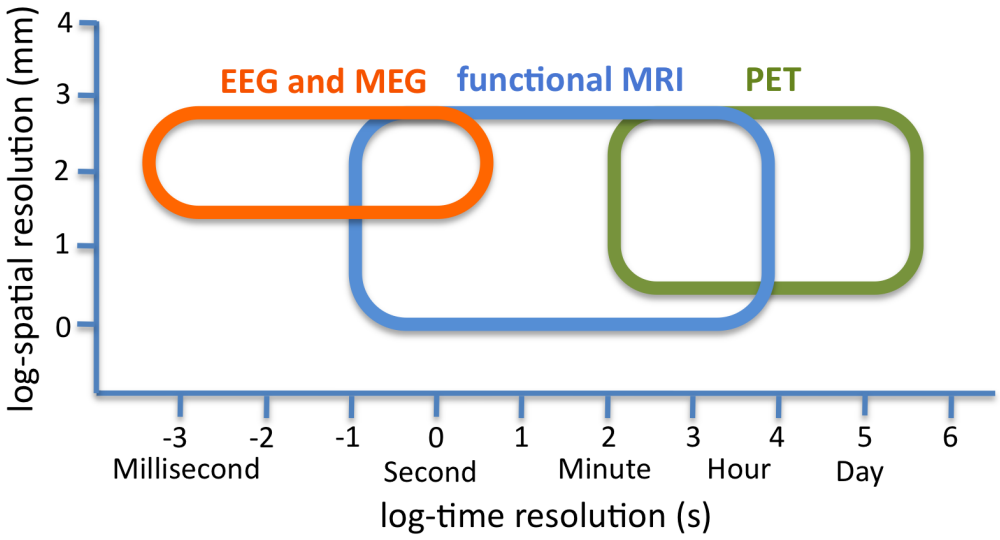

Figure 1.1: Relative temporal and spatial resolution of common brain imaging techniques, such as EEG, MEG, fMRI and PET. As can be seen, EEG and MEG provide the best temporal resolution in general, while fMRI provides the best spatial resolution.

2004].

Human EEG is recorded non-invasively using small electrodes (usually between 32 and 256) that are temporarily fixed to the scalp. Each scalp electrode records electric potentials generated in the tissue beneath it, which contains from 10^7 to 10^9 neurons [Nunez and Cutillo, 1995]. The sum of the electrical activity generated by these cells produces a single current dipole moment in each tissue volume. Tissue volumes close to the scalp surface are much more likely to contribute to extracranial electrical fields, which means the neocortex is, in most cases, the largest contributor to the EEG signal. In addition to its proximity to the EEG sensors, the main neocortex cells, the pyramidal cells, are aligned parallel to each other and normal to the scalp, which facilitates the summation of their electrical activity. In general, any brain structure where cells do not have similar spatial orientation does not create effects that can be macroscopically detected. Activity in sulci is also less reflected in the EEG because dipoles in opposing cortical surfaces within the sulci can cancel out [Niedermeyer and Lopes da Silva, 2004].

Synchronised activity of spatially aligned cells, such as the pyramidal cells, results in the temporal overlap of their extracellular currents, which are induced by the

movement of ions across the cell membrane. Currents from non-phase locked sources, on the other hand, may cancel out. For this reason, EEG signals reflect mainly post-synaptic (dendro-somatic) potentials, which can more easily overlap, both spatially and temporally, than the faster and more focal action potentials (spiking activity) [Nunez and Silberstein, 2000].

Scalp EEG activity shows spontaneous and induced oscillations at a variety of frequencies [Nunez and Cutillo, 1995]. Several of these macroscopic oscillations have characteristic frequency ranges, spatial distributions and are associated with different states of brain functioning (e.g. waking and the various sleep stages) [Nunez and Cutillo, 1995].

The problem of computing the contribution of cortical sources to the scalp EEG potential is called the Forward problem. It needs the specification of a head model, which contains geometric and conductive information about the head volume conductor, and can be relatively straightforward to compute when the head conductivity profile is known (which, in general, it is not). Head models range from the simple assumption that the head is a sphere, or set of multiple spheres, to Boundary Element Methods (BEMs) and Finite Element Methods (FEMs), which allow varying degrees of conductive anisotropy [Baillet et al., 2001b].

The inverse problem is not easy to solve and consists of using the recorded scalp potentials with some constraints (usually assumptions) to find the source distribution that generated the EEG recordings [Nunez, 1990]. The essence of the inverse problem lies in the lack of a unique mathematical solution, and therefore in the infinite number of possible locations and magnitudes of the electric current sources within the brain, even when geometry and conductivity of the different regions is known. Since the inverse problem has no unique solution, any inverse solution depends critically on the chosen constraints, for example, on the number of sources, source distribution, or spatial and temporal smoothness criteria. For a review on solutions to the EEG inverse problem see Baillet et al. [2001a].

Intra-cranial measurements can also be acquired in special cases, such as in the study of epilepsy patients [Tao et al., 2005] or non-human mammals [Logothetis, 2008]. Such intracranial recordings provide measures of cortical dynamics at multiple spatial scales depending on electrode size. Invasive electrodes usually measure Local Field Potentials (LFPs) and Single/Multi-Unit activity (S/MUA). These signals are obtained by filtering the electrode signal using a low-pass filter (cut-off <200 Hz) in the case of LFPs and a high-pass filter (cut-off between 300 and 400 Hz) for MUA. For

this reason, LFPs correspond primarily to weighted averages of synchronised dendrosomatic components of synaptic signals in a neuronal population close to the electrode tip, whilst S/MUA measures the action potentials of a single cell or population of cells, respectively [Logothetis, 2008]. One of the main benefits of such intracranial signals is the preservation of high frequency signals (e.g. gamma activity > 40 Hz) that are attenuated due to spatial smearing at the scalp, in EEG.

In summary, EEG and MEG are the only widely available and non-invasive technologies that provide direct measurements of brain activity with sufficient temporal resolution to follow its fast dynamic changes (10 to 100 millisecond range) (Figure 1.1). However, as mentioned above, the ill-posed EEG inverse problem (also present in MEG) does not allow one to uniquely specify the location of underlying bioelectric activity without additional (prior) information. This fact renders the spatial resolution of EEG and MEG, in general, relatively poor when compared to other techniques, such as fMRI or PET (Figure 1.1).

1.1.2 Multimodal imaging

Given the complementary temporal and spatial characteristics of EEG and fMRI, described in the previous section (1.1.1), the combination of these two techniques should allow one to make inferences about brain function with improved overall spatiotemporal resolution. This section introduces some of the challenges and alternative approaches for integrating EEG and fMRI information, also known as EEG-fMRI fusion. In addition, this section focuses on what can be learnt about the relationship between underlying neuronal activity and hemodynamics (neurovascular coupling) by exploiting the complementary nature of these two measuring techniques.

1.1.2.1 EEG-fMRI integration

Depending on the nature of the experimental question, multi-modality studies with EEG and fMRI rely on datasets that are collected in separate sessions (EEG outside the MR environment) or simultaneously (EEG inside MR scanner). Acquiring the data in separate sessions is the most straightforward procedure because it only needs standard unimodal data acquisition techniques. However, it is suboptimal in problems commonly addressed by EEG-fMRI integration, such as the investigation of resting state or epileptic activity. In addition, habituation effects, variations in the stimulation paradigm, or any other differences between sessions might lead to

differential activity of neuronal networks, which can hinder integration [Rosen et al., 1998; Wagner and Fuchs, 2001]. This includes trial-to-trial variability in response magnitude [Debener et al., 2006].

Simultaneous EEG-fMRI acquisition techniques have been developed to address these issues [Laufs et al., 2008; Ritter and Villringer, 2006]. Advances in hardware, such as the use of MR compatible EEG systems, have been able to minimise the interference of rapidly changing gradient and radio-frequency (RF) electromagnetic fields with the EEG signal. However, despite these advances the signal-to-noise ratio (SNR) of this signal is still significantly lower than in the corresponding unimodal acquisition.

Artifacts that contaminate EEG when inside the MR environment result from the movement of electrical charges (e.g. electrodes) in the static MR field, and from the RF- and gradient switching of the MR scanner [Ritter and Villringer, 2006]. The first type of artifact is mainly dominated by the so-called 'ballistocardiogram' artifact, which is thought to be caused by cardiac-related body and electrode movement due to expansion and contraction of scalp arteries with heart beat. Various post-acquisition methods exist that significantly reduce ballistocardiogram in the EEG. One option is to use template subtraction from individual artifact epochs [Allen et al., 1998], another is to subtract an amplitude adapted dynamic template calculated by sliding average [Kruggel et al., 2000].

Artifacts related to gradient switching and RF pulses typically have high amplitudes in the range of several millivolts (as compared to the EEG signal, which is in the range of tens or hundreds of μV). These artifacts render the EEG signal extremely distorted for the period of MR acquisition. Moreover, because they largely exceed the physiological component of the EEG signal and contain components in the EEG frequency range, elimination of these artifacts is not easy. Effective gradient artifact removal methods include subtraction of a weighted average artifact [Allen et al., 2000] or subtraction of template artifacts adapted to the power spectrum [Sijbers et al., 1999] of individual artifact periods. The particular artifact removal methods used in this thesis are described in Chapters 3 and 5.

Once the EEG data has been artifact-corrected, one can proceed with EEG-fMRI integration using one of three schemes (Figure 1.2): integration through prediction, integration through constraints and integration through fusion. For a recent review of these techniques see Rosa et al. [2010a].

In 'integration through prediction' studies (Figure 1.2, dotted blue line), the objec-

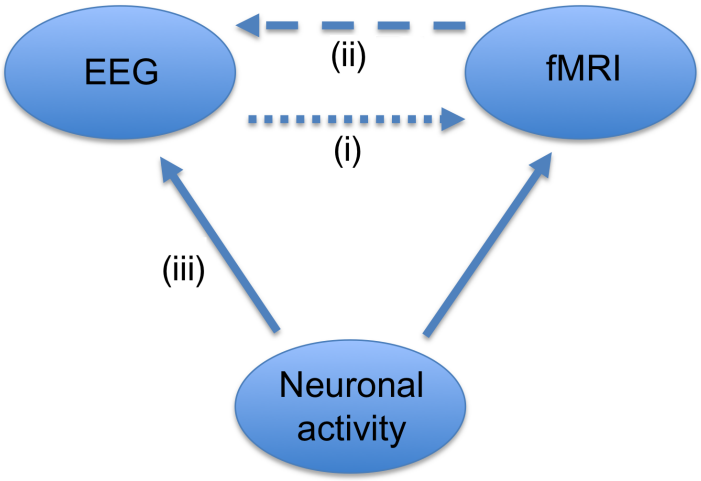

Figure 1.2: Schematic showing the approaches to EEG-fMRI integration. (i) Integration through prediction. (ii) Integration through constraints. (iii) Integration through fusion with forward models [Kilner et al., 2005].

tive is to localise, using fMRI, brain regions whose response is temporally correlated with a given EEG event or feature. However, this integration scheme does not identify regions that are necessarily the generators of the EEG signal, but only those for which BOLD is temporally correlated with the EEG measure used. In other words, one uses EEG time-series as predictors of changes in simultaneously recorded fMRI. The resulting region-specific hemodynamic correlates can then be characterised with high spatial resolution with conventional imaging methodology [Lemieux et al., 2001; Rosa et al., 2010b] (see below for more information on methods for analysing fMRI data, Section 1.2). Typical applications of this approach therefore include hemodynamic correlates of modulations in ongoing oscillatory activity, including epilepsy and sleep. Chapter 3 reviews in more detail some of the main findings obtained using this integration approach. In addition, Chapter 3 uses integration through prediction to investigate the hemodynamic correlates of different features of the EEG frequency spectrum (see also Rosa et al. [2010b]).

The aim of 'integration through constraints' (Fig. 1.2 dashed blue line) is to use fMRI activations to constrain the EEG source reconstruction problem and provide estimates of equivalent dipoles or distributed sources with higher spatial resolution than otherwise possible [Bonmassar et al., 2001]. This is done by apriori positioning dipoles at the location of fMRI activation foci [Vanni et al., 2004], or by penalising

distributed sources whose fMRI-derived activation probability is low [Phillips et al., 2002; Toma et al., 2002]. Although this type of integration has provided good source localisation results, it can be strongly biased when the generators of EEG and BOLD do not overlap [Dale et al., 2000; Liu et al., 1998]. Moreover, EEG and fMRI measures can be sensitive to different kinds of source activity. For example, a region may have a large hemodynamic response contributing to the fMRI signal but the orientations of its cells might not generate a macroscopic potential that can be measured with EEG. The opposite scenario can also happen. For example, changes in neuronal activity captured in the EEG signals (such as frequency, phase or synchronisation changes) might not produce sufficient metabolic activity to generate a BOLD response [Ekstrom et al., 2009]. In these situations, fMRI constraints may therefore considerably distort the resulting EEG source estimates [Liu et al., 2006].

Both integration through prediction and integration through constraints rely on importing information from one modality into the analysis of another, by treating one modality as a predictor or cause of the other, respectively. These approaches are labelled as *asymmetrical*, because they do not analyse the EEG and fMRI data jointly. In contrast, the third approach, 'integration through fusion', is considered *symmetrical* because it depends bilaterally on both datasets [Daunizeau et al., 2007]. For a recent review of symmetrical EEG-fMRI fusion see Valdes-Sosa et al. [2009]. This approach relies on a common generative (forward) model that links the underlying neuronal dynamics of interest to measured hemodynamic and electrical responses (Figure 1.2, solid blue lines). EEG and fMRI data can then be jointly used to invert this model and infer about hidden causes. This approach was first proposed to circumvent certain specific limitations of the asymmetrical approaches (such as the utility of spatial fMRI priors in the source reconstruction problem), but it turned out to be most useful for studying neurovascular coupling. Chapter 5 provides a review of existing integration through fusion approaches and uses one of these frameworks to investigate different biologically plausible neurovascular coupling mechanisms. These mechanisms are described in the following section.

1.1.2.2 Neurovascular coupling

The term 'neurovascular coupling' refers to the relationship between local neuronal activity and subsequent changes in CBF and BOLD signal. Despite the increasing amount of literature in this field this issue is still under intense debate (for a recent

review, see Riera and Sumiyoshi [2010]). This discussion results in part from the fact that neuronal activity within a cortical unit and its relationship to the surrounding vasculature can be described in many different ways and on many different scales. For example, one can focus on the role of post-synaptic potentials versus spiking activity, excitatory versus inhibitory activity, macroscopic oscillations versus membrane potential or even neuronal versus glial cell activity.

Most of our present knowledge about neurovascular coupling comes from animal experiments. These studies usually combine hemodynamic measures such as CBF, with electrical measures such as LFPs and S/MUA. In a pioneering study, Logothetis et al. [2001] found, in monkey visual cortex, that although both LFPs and MUA correlate with the BOLD response, this response could be predicted more accurately from LFPs. This result has been confirmed in awake animals [Goense and Logothetis, 2008]. On the other hand, Rees et al. [2000], Heeger et al. [2000] and Smith et al. [2002] found strong positive correlations between blood flow and spiking activity. However, in Rees et al. [2000] and Heeger et al. [2000] this activity is not actually measured directly but inferred from the coherence of a motion stimulus and a database of previous microelectrode recordings [Geisler and Albrecht, 1997]. More recently, Thomsen et al. [2004], Viswanathan and Freeman [2007] and Rauch et al. [2008] have shown that when synaptic and spiking activity is uncoupled (by drug injection in Thomsen et al. [2004] and Rauch et al. [2008], and using a stimulus that elicits only synaptic activity in Viswanathan and Freeman [2007]), changes in CBF (a predictor of BOLD) do not reflect underlying spiking activity and relate closer to LFPs. Table 1.1 provides a summary of these finding.

This growing body of evidence therefore supports the hypothesis that BOLD signals are more closely coupled to synaptic input and processing activity than to the output spikes of a population of neurons. In addition, this work (Table 1.1) provides support to a growing consensus in which the BOLD signal is thought to result from pre-synaptic activity and the release of neurotransmitters, in particular glutamate [Bonvento et al., 2002]. An increase in pre-synaptic activity and concomitant release of glutamate induces fluctuations in transmembrane potential at the post-synaptic neuron, and these fluctuations are measured with LFPs [Friston, 2008]. At the same time, this activity is also thought to be responsible for triggering the release of vasodilatory agents, such as nitric oxide [Estrada and DeFelipe, 1998; Wang et al., 2005], to the extracellular medium, which induce changes in blood flow and consequently the BOLD response [Bonvento et al., 2002].

The role of interplay between neurons and glial cells in neurovascular coupling has also been emphasised in recent studies (for a recent review see Carmignoto and Gomez-Gonzalo [2010]). These cells are good candidates for generating the vasodilatory agents in response to glutamate release (and perhaps other neurotransmitters). In particular, it is currently accepted that glutamate released by neurons is responsible for increases in intracellular calcium in surrounding astrocytes, which are responsible for triggering the delivery of vasodilatory and constriction agents [Filosa and Blanco, 2007].

When it comes to the human brain, however, most studies address the question of how BOLD relates to underlying neuronal mechanisms only indirectly (by looking at the relationship between (macroscopic) EEG/MEG and fMRI measures, as described below). The number of human studies that investigate particular neurovascular coupling mechanisms, such as synaptic versus spiking activity, is considerably smaller than the number of animal studies. Moreover, these studies come exclusively from neurosurgical patients, whose physiology may be compromised (Table 1.2). Of the few such studies, Mukamel et al. [2005] observe significant correlations between BOLD signals and both synaptic and spiking signals in auditory cortex of epilepsy patients, whilst Ekstrom et al. [2009] found no correlation between BOLD signals and spiking activity in patients' hippocampal area.

These data therefore do not seem to support the existing animal view and there is a clear lack of evidence from healthy human brain. Chapter 5 addresses this issue, by presenting and evaluating a modelling framework that can be used to non-invasively compare different biologically plausible hypotheses for neurovascular coupling. In particular, in Chapter 5 this framework is used to explore the contribution of synaptic and spiking activity to the generation of hemodynamic signals in healthy human visual cortex, with EEG-fMRI data.

The link between neuronal activity and the BOLD response has not only been investigated at a microscopic level, using invasive co-localised recordings, but also at a macroscopic scale using fMRI and EEG/MEG from healthy subjects, as mentioned above. These studies have mainly focused on correlations between BOLD signals and oscillatory EEG power measured in different frequency bands (Table 1.2). The main conclusion of this body of work is that, in general, increases/decreases in power at low frequencies of the EEG spectrum, and decreases/increases at high frequencies, are associated with decreases/increases in the BOLD signal (Table 1.2) [Goldman et al., 2002; Laufs et al., 2003; Moosmann et al., 2003; Rosa et al., 2010b]. However,

questions still remain regarding the exact relationship between BOLD and the frequency content of neuronal activity. In particular, it is still unclear if BOLD depends on the total or relative power of a particular frequency band or combination of bands. A more detailed review of these questions and current findings is provided in Chapter 3. Chapter 3 explores these issues by comparing different frequency-dependent and -independent coupling functions, using simultaneous EEG-fMRI data from healthy subjects.

Even though these macroscopic studies do not directly address the particular biophysical mechanisms responsible for the coupling, they seem to point in the direction of the biological hypothesis constructed from invasively acquired animal evidence (see above): increases in pre-synaptic activity, decrease effective membrane time-constants and result in faster oscillatory dynamics; at the same time more neurotransmitters are released (e.g. glutamate), which lead to increases in BOLD signal [Friston, 2008].

In summary, despite the emerging consensus on how blood flow relates to neuronal activity, there are still issues that need further enlightenment. This thesis addresses two of these issues: how fMRI is related to the frequency profile of EEG (Chapter 3) and how BOLD is related to synaptic and spiking activities in healthy human brain (Chapter 5). Understanding the nature of the link between neuronal activity and BOLD plays a crucial role not only in improving the interpretability of BOLD imaging but also in relating hemodynamic measures to other measures of human brain function.

1.2 Modelling brain function

The aim of this section is to introduce the key models used in imaging neuroscience and how they relate to each other. These models can be quite diverse, ranging from conceptual models of functional anatomy to mathematical models of neuronal and hemodynamics. The brief review presented here is restricted to the statistical and dynamical approaches employed in imaging neuroscience, rather than the much broader field of computational neuroscience [Dayan and Abbott, 2002].

It is common to categorise the existing frameworks into two groups, depending on how they approach the functional organisation of the brain (Figure 1.3) . These two modelling perspectives are: *functional specialisation* and *functional integration* [Friston, 2003a; Friston et al., 2005b]. The aim of the former is to identify *where* evoked

brain responses are expressed, whilst the latter focuses on *how* neuronal responses are caused.

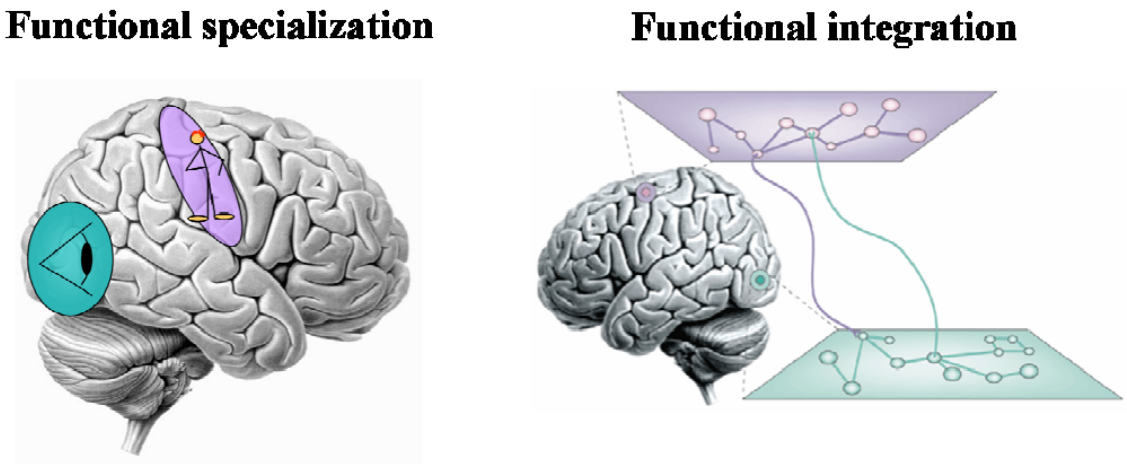

Figure 1.3: Schematic representation of the two different perspectives on functional brain organisation. On the left, functional specialisation aims to identify *where* evoked brain responses are expressed. On the right, functional integration looks at *how* neuronal responses are caused, e.g. how interactions among brain regions are mediated.

The following sections first deal with models of functional specialisation and return to models of functional integration later.

1.2.1 Functional specialisation

From a historical perspective, the identification of a brain region with a specific function has been the central theme in neuroscience. The efforts to attribute function to a particular cortical area, functional localisation, ranged from early anatomical studies to electrical stimulation and brain lesion observation. However, functional localisation *per se* was not easy to demonstrate because it discounted underlying connections and interactions between areas.

More recently, brain research has focused not on attributing a particular function to a particular area, but on identifying features or sub-functions, such as perceptual or motor processing, that can be anatomically segregated within the cortex. This perspective, known as functional specialisation, suggests that experimental manipulation leads to activity changes in, and only in, certain specialised brain areas, without

discounting existing interactions between regions that together enable brain function.

Presently, given the availability of non-invasive imaging techniques, such as fMRI and EEG, functional specialisation studies, or functional 'mapping', typically amount to the production of three-dimensional images of neuronal activation showing which parts of the brain respond to a given cognitive or sensory challenge. This procedure is usually based on some form of statistical parametric mapping (SPM). This framework is described below.

1.2.1.1 Statistical Parametric Mapping

SPM is a modelling framework used to test hypotheses about regionally specific effects in the brain, also known as brain activations [Friston et al., 1995]. Statistical Parametric Maps (SPMs) are images constructed with statistical values that, under the null hypothesis, are distributed according to a known probability density, usually the Student's t or F-distribution (t- and F-maps, respectively) (Figure 1.5).

SPM was introduced almost two decades ago [Friston et al., 1995], and is undoubtedly one of the most used techniques for fMRI data analysis. The idea behind this framework is quite simple: the data from each and every voxel is analysed independently using a general linear model (GLM) and standard univariate (parametric) statistical tests. The resulting voxel-wise statistics are assembled into an image and interpreted as continuous statistical processes, by referring to the probabilistic behaviour of random fields, modelled by random field theory (RFT) (see below) [Adler, 2006; Worsley et al., 1992, 1996].

The GLM, also known as analysis of (co)variance or multiple linear regression, is an equation that expresses the observed response variable in terms of a linear combination of explanatory variables. The matrix that contains the explanatory variables is called the design matrix. Each column of the design matrix corresponds to an effect one has built into the experiment or that may confound the results. The columns are also referred to as covariates or regressors. Each regressor has an associated unknown parameter and its relative contribution to explain the data is assessed using standard maximum likelihood [Christensen, 2002].

For fMRI data, the independent variables representing the experimental effect are convolved with a hemodynamic response function (HRF) [Boynton et al., 1996] (Figure 1.4). This function accounts for the fact that BOLD signals are a delayed and dispersed version of the neuronal response, furnishing a better statistical model of

the data. It is also common to convolve the regressors in the design matrix with the HRF's first temporal and dispersion derivative (Figure 1.4). These two derivatives allow for the time and duration of the peak of the hemodynamic response to vary from voxel to voxel and subject to subject. The temporal derivative, for example, allows for peak responses that are approximately one second earlier or later than is usual, whilst a positive/negative estimate for the dispersion derivative implies a less/more dispersed response than usual (Figure 1.4).

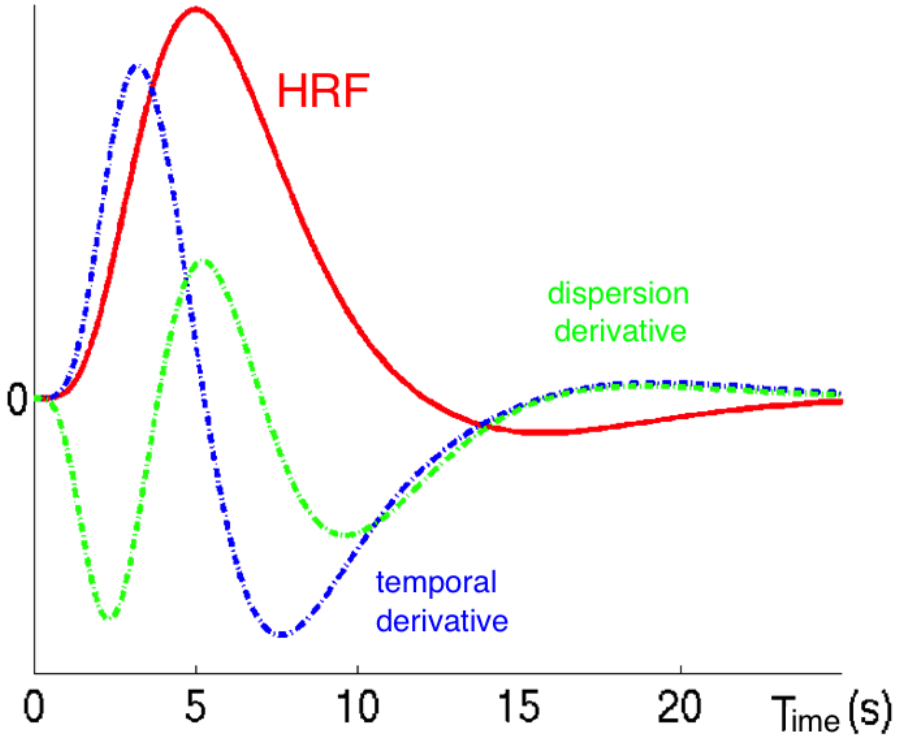

Figure 1.4: HRF (red) and its time (blue) and dispersion (green) derivatives. These two derivatives allow for the time and duration of the peak of the hemodynamic response to vary from voxel to voxel and subject to subject. The temporal derivative, for example, allows for peak responses that are approximately one second earlier or later than is usual, whilst a positive/negative estimate for the dispersion derivative implies a less/more dispersed response than usual.

The SPM framework can also be used to create scalp-maps, time-frequency images, and volumetric 3D source reconstruction images using EEG/MEG data [Litvak et al., 2011]. These statistical maps are again based on the GLM and corrected for multiple comparisons using RFT [Kilner and Friston, 2010], as described in the

following section. In addition, the SPM approach can also be used with structural data, acquired with MRI, to look for voxel-wise differences in the local density of gray matter between groups of subjects. This technique is known as Voxel-Based Morphometry (VBM) [Ashburner and Friston, 2000].

1.2.1.2 Classical and Bayesian inference

Classical inferences about the parameter estimates are made using their estimated variance, allowing one to test the null hypothesis, that all the estimates are zero, using the F-statistic or that some particular linear combination (e.g. a subtraction) of the estimates is zero, using a t-test [Poline et al., 1997]. The t-statistic obtains by dividing a 'contrast' vector of the ensuing parameter estimates by the standard error of that contrast [Poline, 2003].

However, in classical inference, without any apriori anatomical hypothesis, a correction for multiple comparisons over the volume analysed is necessary to ensure that the probability of rejecting the null hypothesis incorrectly (false positives) is maintained at a small rate. RFT provides a way to accommodate this problem by adjusting the voxel-wise p-values, whilst taking into account the fact that neighbouring voxels are not independent by virtue of continuity in the original data [Adler, 2006; Worsley et al., 1992, 1996]. RFT plays the same role for continuous statistical fields (in this case, images), as the Bonferroni correction for discrete statistical tests, therefore being less conservative than the latter. In other words, RFT deals with both the univariate probabilistic characteristics of an SPM and any non-stationary spatial covariance structure in the data. As a result, any topological features of the SPM that are 'unlikely' to occur by chance, are interpreted as regionally specific effects attributable to the experimental manipulation.

However, if one has an apriori anatomical hypothesis about particular regions in the brain that are likely to activate with the stimulus, e.g. the Brocas area during word generation, one may chose to correct for multiple comparisons only within these regions. This procedure is called small volume correction (SVC) and is less conservative than taking into account the whole brain. In Chapter 3, SVC is used to correct the p-values of voxels within the primary visual cortex, in a visual stimulation task.

An alternative approach to classical inference is to use Bayesian inference [Friston et al., 2002b]. This approach is based upon the posterior distribution of the effect,

i.e activation, given the data. Bayesian inference relies on the specification of a prior probability distribution, which comprises knowledge or beliefs about the effect that have been obtained before observing the data. After observing the data, these priors are updated into the posterior distribution. A common way to summarise this posterior is to compute the probability that the effect of interest exceeds some threshold. By computing this probability for each voxel, one can again assemble the voxel-wise statistics into a three dimensional image (Figure 1.5), in this case known as a posterior probability map (PPM) [Friston and Penny, 2003].

The motivation for using Bayesian inference is that it has high face validity [Gelman et al., 1995]. This is because the inference is about an effect being greater than some specified size that has some meaning in relation to the underlying neurophysiology. This contrasts with classical inference, in which the inference is about the effect being significantly different from zero. The problem with classical inference is that, with sufficient data or sensitivity, trivial departures from the null hypothesis can be declared significant.

Another advantage of using Bayesian inference in neuroimaging is the fact that it does not contend with the multiple comparisons problem. The probability that activation has occurred, given the data, at any particular voxel is the same, irrespective of whether one has analysed that voxel or the entire brain. Because there is no need for false positive rate correction, PPMs can be relatively more powerful than SPMs [Friston and Penny, 2003]. However, see Woolrich [2011] for an alternative view.

In this thesis, the SPM framework is used in Chapter 3 to compare different neurovascular coupling functions, instantiated as linear models. These functions (GLM regressors) are constructed using different features of the EEG power spectrum, and are then voxel-wise regressed onto the fMRI data. Inferences are based on F-maps. In addition, Chapter 4 presents a new method for comparing nested and non-nested GLMs based on PPMs for inference on model space. The full description of this work can be found in Chapters 3 and 4, respectively.

1.2.1.3 Biophysical models

In the modelling framework described in the previous sections, regionally specific responses were modelled as linear mixtures of designed changes in explanatory variables. Although useful, these models do not allow one to make inferences about hidden quantities, such as volume and deoxyhemoglobin concentrations, that are not

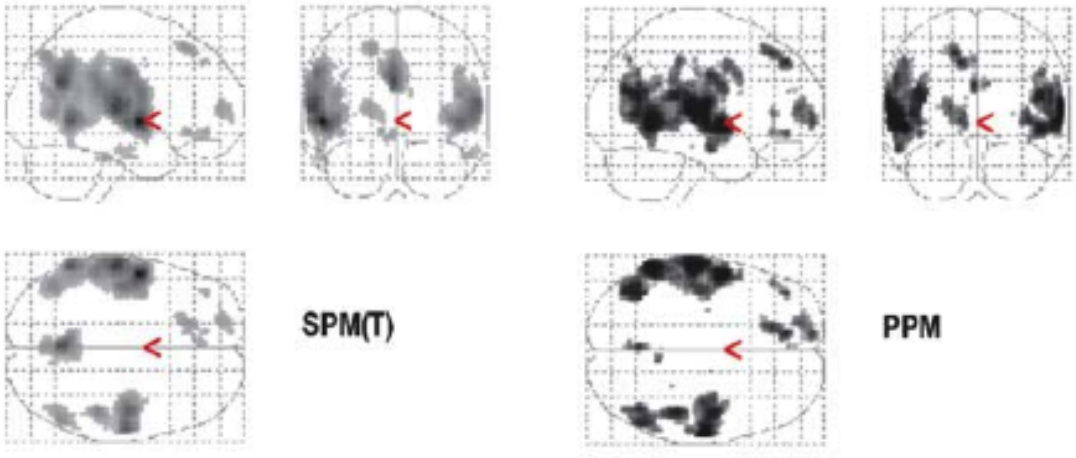

Figure 1.5: Illustration of classical and Bayesian inference for a functional mapping study of word generation. (a) Classical SPM of the t-statistic for a contrast reflecting the difference between word shadowing and word generation. This SPM has been thresholded at p-value < 0.05, corrected using RFT. (b) PPM for the same contrast, using an activation threshold of 2.2 and a confidence of 95%. Figure adapted from Friston and Penny [2003].

actually observed.

To be able to infer about the hidden biophysical mechanisms that give rise to the observed data, one needs to finesse the modelling framework with a generative model of how the data are caused. This generative approach to modelling brain responses has a much more direct connection with underlying physiology and rests upon an understanding of the underlying system [Horwitz et al., 2000].

From a mathematical point of view, these models usually comprise an input-state-output dynamical system that expresses the rate of change of the states as a parameterised function of the states and inputs (Figure 1.6). Typically, the inputs correspond to designed experimental effects (e.g. the stimulus function in fMRI). The objective is then to make inferences about the causes and learn the parameters.

However, in contrast to the simpler GLMs, these models are less straightforward to invert, due to their extra complexity. One of the procedures commonly used in functional imaging, is to use Bayesian inference and posit an approximate posterior density that is consistent with the generative model, and which can be parameterised. Inference then proceeds by estimating the moments (e.g. expectation) of this density, within a well-known optimisation scheme, such as the Expectation-Maximisation

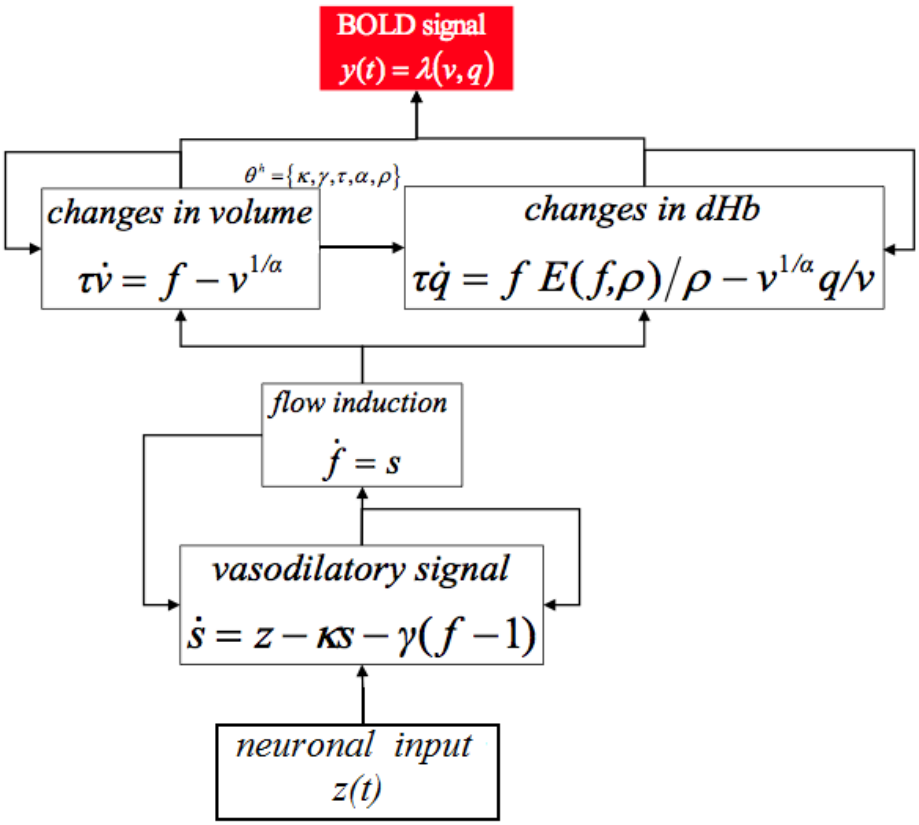

Figure 1.6: Extended Balloon model [Friston, 2002b]: these equations model the hemodynamic response to neuronal activity, z, for a single region. Neuronal activity induces a vasodilatory signal s that increases blood flow f. Increases in blood flow cause changes in volume and deoxyhemoglobin (v and q). These two hemodynamic states enter the output nonlinearity, λ, to give the observed BOLD response y. The hemodynamic parameters, $\theta_{_H} = \{\kappa, \gamma, \tau, \alpha, \rho\}$, comprise the rate constant of the vasodilatory signal decay, the rate constant for autoregulatory feedback by blood flow, transit time, Grubb's vessel stiffness exponent, and the resting oxygen extraction fraction, respectively. A more detailed description of the model and parameters is presented in Chapters 5 and 6 with a slightly different notation: $\theta_{_H} = \{\tau_s = 1/\kappa, \tau_f = 1/\gamma, \tau_0 = \tau, \alpha, E_0 = \rho\}$.

(EM) algorithm [Friston, 2002b]. This procedure is described in more detail in the next chapter (Chapter 2).

One example of a generative model commonly used in neuroimaging is the Balloon model for fMRI data (Figure 1.6). Instead of using a convolution model to emulate the hemodynamic response, as is done in GLM analyses (Section 1.2.1.1), one can model this response using a biophysically grounded model. The Balloon model, originally proposed by Buxton et al. [1998] and extended by Friston et al. [2000], comprises a set of ordinary differential equations that model how changes in blood flow, evoked by neuronal activity, are transformed into the observed BOLD response (Figure 1.7). This system of equations (Figure 1.6) is driven by neuronal activity, z, which first causes an increase in a vasodilatory signal, s, that is subject to auto-regulatory feedback. Inflow, f responds in proportion to this signal with concomitant changes in blood volume ν and deoxyhemoglobin content q (Figure 1.7). This model is used in Chapter 5 as part of a biophysical framework for modelling both EEG and fMRI signals, whilst in Chapter 6 this model is used to make inferences about effective connectivity, as described in the following Section 1.2.2.

Another example of biophysically motivated frameworks are the neural mass models (NMMs), which model the generation of electrophysiological data, as measured with EEG/MEG [David and Friston, 2003; Jansen and Rit, 1995; Sotero et al., 2007]. A good introduction to NNMs is given in Grimbert and Faugeras [2006]. For further background on these dynamical neural network models the reader is referred to Wilsom [1999] and Ermentrout and Terman [2010].

NMMs can be viewed as a special case of ensemble density models [Deco et al., 2008]. Ensemble models describe the dynamics of large populations of neurons, where single neurons (e.g. pyramidal cells, spiny stellate cells and interneurons) are characterised by a number of states, such as post-synaptic membrane potential and firing rate. In NMMs, the ensemble density of these neuronal states is summarised with a single number representing mean activity [Marreiros et al., 2009]. Assuming the equilibrium density has a point mass (i.e., a delta function), one can reduce the density dynamics to the location of that mass. This reduces the model to a set of non-linear differential equations describing the evolution of this mode. Again, Chapter 5 presents a biophysical framework for modelling both EEG and fMRI signals, by means of combining NMMs with the Balloon model described above [Riera et al., 2007].

NMMs are also used in brain connectivity studies, described in the following

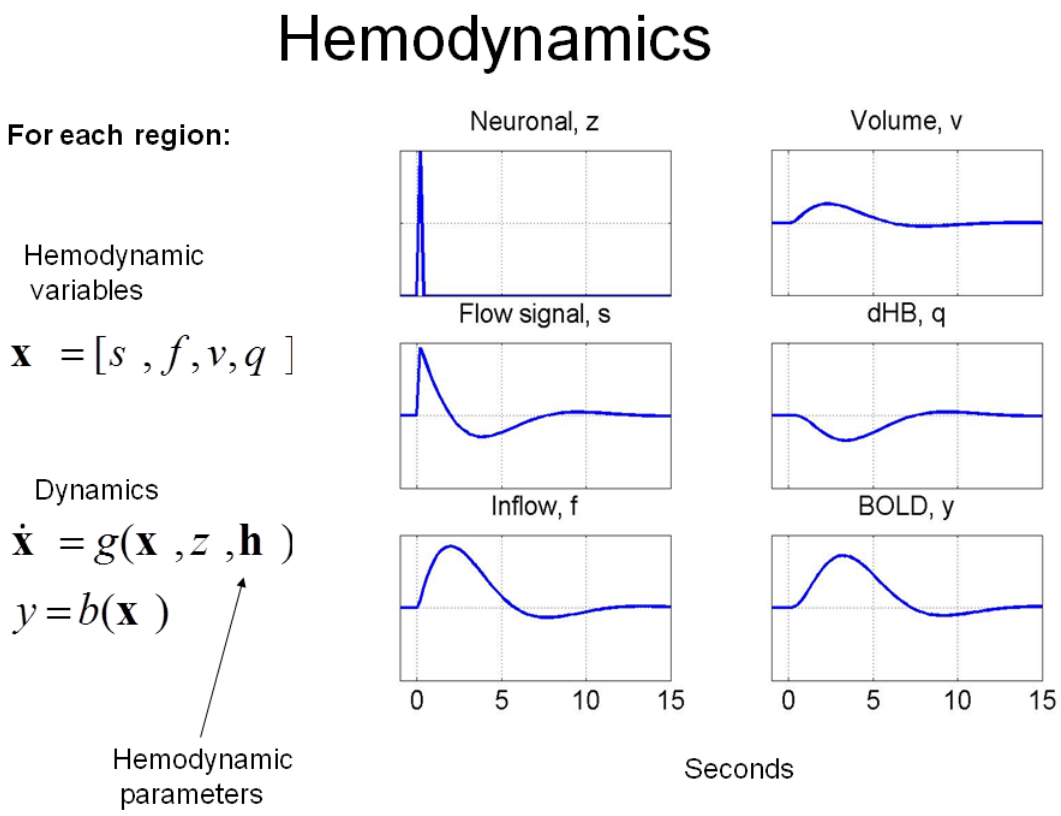

Figure 1.7: Hemodynamics generated using the extended Balloon model [Friston, 2002b] in response to a brief stimulus. These signals correspond to the equations plotted in Figure 1.6. A more detailed description of the model and parameters is presented in Chapters 5 and 6.

section (1.2.2), using EEG and MEG data. For a review see [Kiebel et al., 2009].

So far, the models introduced here concern responses at a single brain region or voxel. The following section retains the same biophysical perspective on modelling brain function, but in the context of distributed responses and functional integration.

1.2.2 Functional integration

Functional specialisation based on SPM has proved to be a very powerful methodology for visualising regionally specific effects. However, understanding macroscopic brain function ultimately requires knowing not only about the behaviour of specific brain areas but also about how these areas interact with each other.

Functional integration refers to the study of interactions among specialised neuronal populations and how these interactions depend upon the sensorimotor or cognitive context [Friston, 2002a]. By contrast to functional specialisation and SPM, the study of functional integration has relied on a wide range of statistical techniques. These techniques include both unsupervised (e.g., independent component analysis (ICA) [Brown and Marsden, 2001]) and supervised methods (e.g., support vector machine (SVM) [Mourao-Miranda et al., 2005]). In addition, they also include models that seek to directly measure 'causal' connectivity based on either static statistical constraints (e.g., structural equation modelling (SEM) [McIntosh and Gonzalez-Lima, 1994]) or dynamic and more bio-physically motivated assumptions (e.g., dynamic causal modelling (DCM) [Friston et al., 2003]).

Given the variety of methods, it is common to classify these methods into two categories within functional integration: *functional connectivity* and *effective connectivity*. Functional connectivity is defined as any statistical dependencies (e.g. correlations or coherence) between remote neurophysiological events, independent of how these dependencies are mediated. This is typically a data driven approach that does not allow one to infer about the directionality of the dependencies [Rogers et al., 2007]. Effective connectivity, on the other hand, refers explicitly to the influence that one neuronal system exerts over another, either at a synaptic (i.e. synaptic efficacy) or population level, and therefore depends upon a model of the interactions, i.e. it is a model driven approach [Stephan and Friston, 2010].

The following section focuses on one of the most used effective connectivity approaches in imaging neuroscience, dynamic causal modelling.

1.2.2.1 Dynamic causal modelling

DCM is a mathematical framework to estimate, and make inferences about, the coupling among brain areas and how this coupling is influenced by changes in experimental context [Friston et al., 2003].

The idea behind DCM is to construct a reasonably realistic model of interacting cortical regions or nodes. To this neuronal network model is added a forward model of how neuronal activity in each cortical region is transformed into a measured response, such as electromagnetic or hemodynamic signals. This generative framework enables the parameters of the neuronal model (i.e. effective connectivity) to be estimated from observed data.

To date, DCM has been applied to fMRI, EEG and MEG [Kiebel et al., 2009], as well as LFP data [Moran et al., 2009]. This section focuses on DCM for fMRI, as chapter 6 of this thesis investigates a recently developed model selection criterion in this context.

In terms of the mathematics, DCM for fMRI comprises a deterministic bilinear model, or set of bilinear deterministic differential equations (for two-state, non-linear and stochastic DCMs see Stephan et al. [2008], Marreiros et al. [2008b] and Friston et al. [2010a], respectively) that describes the dynamics at the neuronal level. On top of this framework, is the extended Balloon model (as mentioned in the previous section) for the hemodynamic level (Figure 1.8). The full description of the model equations can be found in Chapter 6.

By using a bilinear approximation, to the systems equations of motion, the parameters of the implicit causal model reduce to three sets. These comprise parameters that: mediate the influence of extrinsic inputs on the states; mediate regional coupling among the states; and [bilinear] parameters that allow the inputs to modulate that coupling.

In stochastic DCMs for fMRI the bilinear deterministic equations are replaced by random differential equations and the bilinear (modulatory) and exogenous inputs are discarded (for further details on these models see Friston et al. [2010a]). This method is particularly suitable to study resting-state effective connectivity, and models endogenous fluctuations that cannot be assessed using deterministic DCMs [Friston et al., 2010a].

In current implementations of DCM, independent of modality, the model parameters are estimated from the data using Bayesian methods [Friston et al., 2007a], and

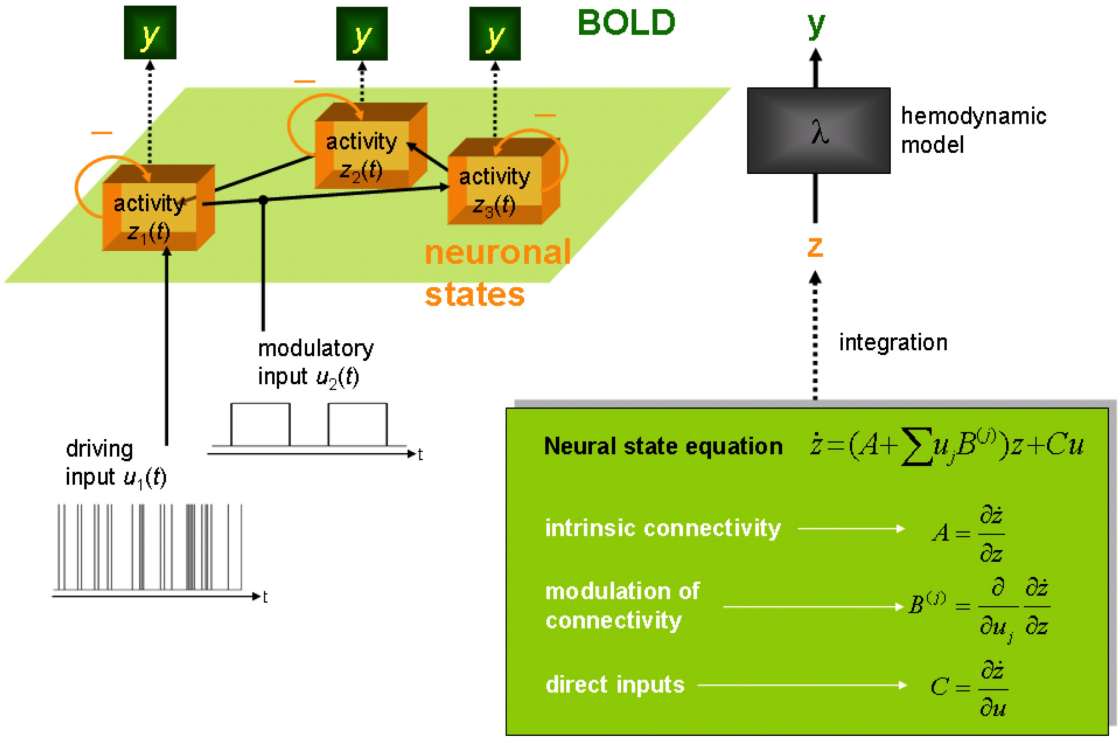

Figure 1.8: Schematic illustrating the concepts underlying dynamic causal modelling. In DCM for fMRI, a set of bilinear deterministic differential equations describes the dynamics at the neuronal level (neuronal states), where the dot notation, $\dot{z}$, denotes a temporal derivative. The extended Balloon model is then used to model the hemodynamic level. Inputs to the system can be of two types: driving inputs, u_1, or modulatory inputs, u_2. The full description of the model equations can be found in Chapter 6.

different models are compared using the model evidence [Penny et al., 2004]. These procedures are described in Chapter 2.

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + u_2 \begin{bmatrix} 0 & 0 \\ b_{21} & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} c \\ 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

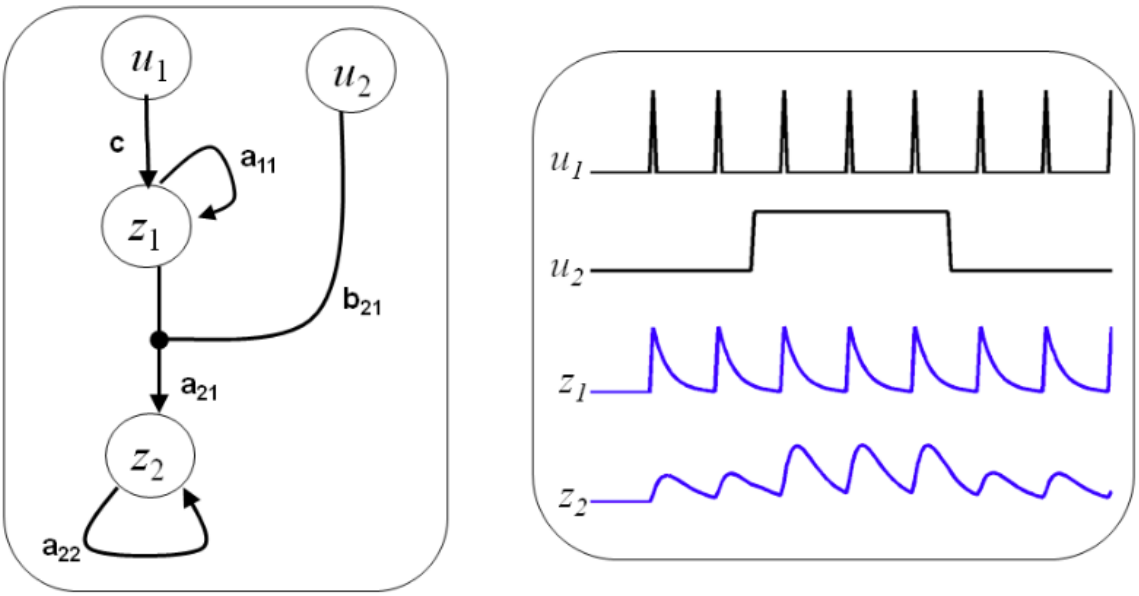

Figure 1.9: DCM with two regions (z_1 and z_2) and two inputs (u_1 and u_2). The equation on top of the figure is the equation for a deterministic bilinear DCM, Eq. 6.1, described in Chapter 6, corresponding to the network on the left. The time-series of the two regions generated by this equation, under the influence of driving input u_1 and modulatory input u_2 are plotted on the right. The full description of the model equations can be found in Chapter 6.

In DCM, an experiment is regarded as a designed perturbation of neuronal dynamics distributed throughout a system of coupled anatomical nodes (Figure 1.9). From this perspective, DCM uses the same experimental design principles to induce region-specific interactions used in conventional region-specific activation studies (Section 1.2.1.1). In fact, the convolution model, used in SPM analyses of fMRI time-series, is a special case of DCM that arises when the coupling among regions is not taken into account. In DCM, the explanatory variables that comprise the design matrix in SPM become the inputs, whilst the parameters become measures of effective connectivity [Friston, 2003b].

Although DCM can be viewed as a generalisation of the GLM, it can also be seen as an attempt to introduce more plausible biophysical models of how neuronal dynamics respond to inputs and produce measured responses. This reflects the growing appreciation of the role that biophysically grounded models have in understanding brain function [Horwitz et al., 2000].

DCM has gradually become part of mainstream neuroimaging analysis techniques and has been applied to a wide range of domains, including: language [Leff et al., 2008; Noppeney et al., 2008; Schofield et al., 2009], motor processes [Grefkes et al., 2008; Grol et al., 2007], vision and visual attention [Fairhall and Ishai, 2007; Haynes et al., 2005], memory [Smith et al., 2006], perceptual decision making [Summerfield and Koechlin, 2008; Summerfield et al., 2006], and learning [den Ouden et al., 2009; Garrido et al., 2009].

Chapter 6 describes and evaluates a new method for comparing large numbers of DCMs using fMRI data from an attention to visual motion study.

1.3 Conclusion

This chapter has reviewed some key imaging and modelling techniques currently used in brain research. It has also focused on how two of these techniques, fMRI and EEG, can be combined for multimodal imaging. This multimodal approach allows one to investigate neurovascular coupling, as described above. From the modelling perspective, this chapter discussed ways of assessing specialisation and integration in the brain. General linear models based on convolution models can be finessed with biophysical constraints to formulate explicit forward models of neuronal activity and hemodynamics. The latter can then be used to assess interactions among distributed cortical areas and make inferences about coupling at the neuronal level. The following chapter will focus on how to invert and compare the models presented here.

Table 1.1: Main findings of animal studies on neurovascular coupling. S/MUA refers to single/multi-unit activity. CBF refers to cerebral blood flow; TO_2 to tissue oxygenation concentration; OHb, dHb, CBV to oxy and deoxy-Hemoglobin and cerebral blood volume, respectively.

Reference	Paradigm	Main findings	Brain regions	Species	Signals
Logothetis et al. [2001]	Visual (rotating checker-board)	LFP (40-130Hz) better predictor of BOLD than MUA (300-1.5kHz)	V1	Monkey	BOLD, LFP, MUA
Goense and Logothetis [2008]	Visual (rotating checker-board)	BOLD's variance best explained by LFP (20-60 Hz)	V1	Monkey (awake)	BOLD, LFP, MUA, SUA
Rees et al. [2000]	Visual (moving dots; changing coherence)	BOLD contrast in human V5 is proportional to SUA in monkey V5	V5	Monkey and Human	BOLD, SUA
Heeger et al. [2000]	Visual (changing contrast)	BOLD in human V1 is proportional to SUA in monkey V1	V1	Monkey and Human	BOLD, SUA
Thomsen et al. [2004]	Resting-state	Drug induced increase in Purkinje cell spike activity was not sufficient to raise blood flow above baseline	Cerebellum	Rat	CBF, SUA
Viswanathan and Freeman [2007]	Visual (sine-wave gratings, 1-20Hz)	Correlation between BOLD and LFPs in the absence of spiking activity (suppressed by the stimulus)	V1	Cat	LFP, MUA, TO_2

Table 1.1 – continued from previous page

Rauch et al. [2008]	Visual (rotating checker-board)	Injected neuromodulator BP554 induces hyperpolarization of efferent membrane, reducing MUA (800-3k Hz) without affecting either LFP (24-90 Hz) or BOLD activity	V1	Monkey	BOLD, LFP, MUA
Kayser et al. [2004]	Visual (sinewave gratings, natural movies and pink pixel noise)	Agreement between BOLD and LFP (in terms of % of recording sites) depends on LFP frequency. Best agreement between 20 and 50 Hz. Poorer agreement for MUA	Visual cortex (17,18,19 and 21a)	Cat	BOLD, LFP, MUA
Niessing et al. [2005]	Visual	BOLD correlates better with gamma-band LFP	Visual cortex	Cat	BOLD, LFP, MUA
Maier et al. [2008]	Perceptual suppresion	Only BOLD and low-Hz LFP (not high-Hz LFP or spikes) significantly decreased during perceptual suppression	V1	Monkey (awake)	BOLD, LFP, Spikes
Boorman et al. [2010]	Whisker pad stimulation	Deep layer negative BOLD, adjacent to layers of positive BOLD, associated with reductions in MUA	Somato-sensory cortex	Rat	BOLD, LFP, MUA, OHb, dHb, CBV

Table 1.1 – continued from previous page

Lee et al. [2010]	Optical stimulus	Negative BOLD signal caused by optically driven genetically modified inhibitory cells	Motor cortex	Rat	Opto-genetics

Table 1.2: Main findings of human studies on neurovascular coupling. ECoG refers to Electrocorticography.

Reference	Paradigm	Main findings	Brain regions	Species	Signals
Mukamel et al. [2005]	Movie segment	Significant correlation between patients' predicted BOLD signals from SUA and signals measured in healthy subjects	Auditory cortex	Human (patients)	BOLD, LFP, SUA
Ekstrom et al. [2009]	Spatial navigation in virtual environment	Correlation between the BOLD signal and theta-band activity; no significant correlation with MUA/SUA	Hippocampal areas	Human (patients)	BOLD, LFP, MUA, SUA
Goldman et al. [2002]; Laufs et al. [2006a]; Moosmann et al. [2003]	Resting-state	Reductions in alpha power correlate with increases in BOLD	Occipital cortex	Human (healthy)	BOLD, EEG
Lachaux et al. [2007]	Semantic decision task	Close spatial correspondence between BOLD activation regions and gamma-ECoG sites	Temporal and sulcal cortex and insula	Human (patients)	BOLD, ECoG
Rosa et al. [2010c]	Visual (flickering checkerboard 4-60Hz)	Root-mean squared frequency explains more BOLD activity than the total spectral power or any linear combination of frequency-bands	Visual cortex	Human (healthy)	BOLD, EEG

Table 1.2 – continued from previous page

Nir et al. [2007]	Movie segments	Gamma-LFP coupled well to BOLD; coupling for SUA highly variable	Auditory cortex	Human (patients)	BOLD, LFP, SUA
He et al. [2008]	Wakefulness (AW), slow-wave and rapid-eye-movement sleep (REM)	State-invariant significant structural correlation between BOLD and slow cortical potentials (<4 Hz). Gamma band potentials only correlate with BOLD during AW and REM	Sensorimotor cortex	Human (patients)	BOLD, ECoG

Chapter 2

Model selection methods

Chapter 1 presented an overview of some of the most commonly used imaging and modelling techniques for investigating human brain function. Given these techniques the next step in learning about the brain typically entails comparing competing hypothesis, or models, such as general linear or biophysical models, for explaining the observed data. This chapter therefore focuses on model selection methods used in neuroimaging, and with which the Results in Chapters 3 to 6 have been obtained. It begins by introducing model inference and its two possible approaches: the Classical and Bayesian approach. The focus here is on the Bayesian approach, since it is the most used throughout this thesis. This chapter then proceeds by reviewing Bayesian model inversion and selection, and, finally, it ends with a description of how inferences can be taken from single subject to group level.

2.1 Inference on model space

Model comparison is a fundamental tool in any scientific discipline. In other words, the process of learning about nature is based on evaluating different hypotheses about how the world works, expressed as mathematical equations in light of experimental observations.

Model comparison is therefore an instrument for making inferences about unobserved processes based on observations. If the experimental data clearly supports one model over several others, this entails strong support to the corresponding hypothesis (among those considered), and the process that most likely generated the data can be inferred (Figure 2.1).

Given the space of all models being compared, model space M, the variable m here indexes members of that space. The models might be GLMs, where m indexes different design matrices (e.g. different transfer functions in Chapter 3 and information processing models in Chapter 4), or biophysical models such as the models used in Chapter 5, where m indexes different neurovascular coupling mechanisms. In DCM (Chapter 6), m indexes different networks, i.e. different connectivity and input patterns.

Model comparison can be implemented using classical or Bayesian inference. The following sections provide a brief summary of these two approaches.

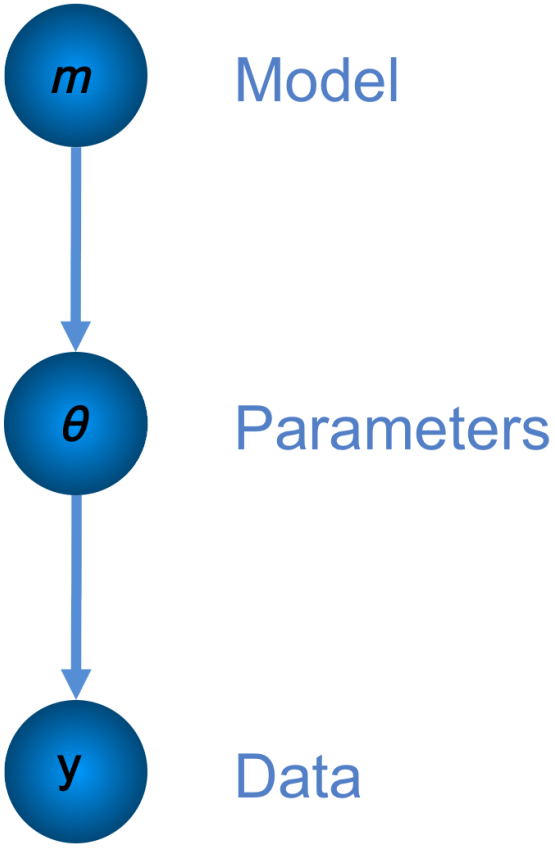

Figure 2.1: Hierarchical generative model: the model m, is part of the hierarchical process of generating data. In Bayesian inference, first a model is chosen from the prior distribution over models, $p(m)$, then its parameters are generated from the parameter prior, $p(\theta|m)$, according to model m, and finally data are generated from the likelihood $p(y|\theta, m)$.

2.1.1 Classical inference

The dominant paradigm is to generate a null hypothesis (typically a simpler and less biologically meaningful hypothesis) and ask whether this hypothesis can be rejected after observing the data. Rejection occurs when a test statistic generated from observed data falls below an arbitrary probability threshold (usually a p-value < 0.05), which is interpreted as strong evidence in favour of a biologically more meaningful alternative hypothesis. Hence, the actual hypothesis of interest (the alternative hypothesis) is accepted only in the sense that the null hypothesis is rejected.

One can then use a series of null hypothesis tests, such as likelihood ratio tests or F-tests, to compare pairs of models from among the candidate set. In Chapter 3, F-tests are employed to make inferences on competing transfer functions from EEG to fMRI, instantiated as GLMs. This procedure tests the additional variance explained by including a set of additional regressors of interest in the model, which in this case represent a particular neurovascular coupling hypothesis.

Even though this approach is a mainstay in scientific research, it is restricted to nested models (i.e. the simpler model must be a special case of the more complex model). In other words, to infer that changes in an experimental factor led to significant effects one compares two models, one with that factor and one without. Such tests cannot therefore be used to quantify relative support for various models. Whilst sufficient in some cases, such as for analysis of variance (ANOVA), it is suboptimal in other domains. The following section describes an alternative approach.

2.1.2 Bayesian inference

In contrast to classical inference, the Bayesian approach offers a way to draw inferences from a set of multiple competing hypotheses, without using the traditional null hypothesis testing procedure.

Bayesian model selection (BMS) therefore provides several advantages over the classical approach [Wasserman, 2000]. One of the main advantages is the fact that BMS is not limited to comparing nested models. BMS allows models to be ranked and weighted, thereby providing a quantitative measure of relative support for each competing hypothesis, including the null hypothesis. In addition, BMS can be used to identify the single best model that most likely generated the observed data, but it can also be used to make inferences based on the weighted evidence from the set of competing models. The latter is known as Bayesian model averaging (BMA) [Hoeting

et al., 1999], and allows incorporation of model uncertainty into inference about the parameters.

BMS has been extensively used in neuroimaging, specially with DCMs [Friston and Penny, 2003; Penny et al., 2005] and EEG source reconstruction models [Trujillo-Barreto et al., 2004]. In this thesis, BMS is used in Chapter 4 to construct posterior model probability maps. In Chapter 5, BMS is used to compare biophysically motivated hypotheses for neurovascular coupling, while in Chapter 6, BMS is used to compare DCMs.

The following sections describe Bayesian model inversion and selection in more detail, including inferences at the single subject and group level.

2.2 Bayesian model inversion

2.2.1 Posterior probabilities

In Bayesian inference, prior beliefs about parameters, θ, of model m are quantified by the prior density, $p(\theta|m)$. Inference on the parameters, θ, after observing data, y, is based on the posterior density $p(\theta|y,m)$. These densities are related through Bayes' rule:

$$p(\theta|y,m) = \frac{p(y|\theta,m)p(\theta|m)}{p(y|m)}, \tag{2.1}$$

where $p(y|\theta,m)$ is the probability of the data (likelihood) conditioned upon the model and its parameters. The normalisation factor, $p(y|m)$, is called the model evidence and plays a central role in model selection (see below). The *Maximum a Posteriori* (MAP) estimate of the model parameters is simply:

$$\theta_{MAP} = \arg\max_{\theta} p(\theta|y,m). \tag{2.2}$$

The posterior density is an optimal combination of prior knowledge and new observations, and provides a complete description of uncertainty about the model parameters. Generally, the choice of priors reflects either empirical knowledge (e.g., previous measurements) or formal considerations (e.g., biological or physical constraints).

The following sections describe Bayesian model inversion for linear and non-linear models.

2.2.2 Linear models

This section describes how the posterior distribution of the parameters of a linear model of the type described in Section 1.2.1.1 and used in Chapters 3 and 4, can be estimated from the data, y. In what follows, $N(\mu, C)$ denotes a multivariate normal distribution with mean μ and covariance C.

Given a linear model $y = g(\theta) + \epsilon = X\theta + \epsilon$, where X is the design matrix, under Gaussian assumptions about the error, $\epsilon \sim N(0, C_\epsilon)$, the likelihood, $p(y|\theta) = N(X\theta, C_\epsilon)$, and priors, $p(\theta) = N(\eta, \Sigma)$, can be written as:

$$\begin{aligned} p(y|\theta) &\propto \exp\{-\frac{1}{2}(y - X\theta)^T C_\epsilon^{-1}(y - X\theta)\} \\ p(\theta) &\propto \exp\{-\frac{1}{2}(\theta - \eta)^T \Sigma^{-1}(\theta - \eta)\}, \end{aligned} \tag{2.3}$$

where the dependency on m was dropped in order to simplify the notation. Using Eq. 2.3 and Eq. 2.1, the posterior density also has a Gaussian form, $p(\theta|y) = N(\mu, C)$:

$$p(\theta|y) \propto \exp\{-\frac{1}{2}(\theta - \mu)^T C^{-1}(\theta - \mu)\}, \tag{2.4}$$

and the posterior means and precisions, $P = C^{-1}$, are given by:

$$\begin{aligned} P &= X^T C_\epsilon^{-1} X + \Sigma^{-1} \\ \mu &= C(X^T C_\epsilon^{-1} y + \Sigma^{-1}\eta). \end{aligned} \tag{2.5}$$

See Bishop [2006] for a derivation. To compute the posterior moments, Eq. 2.5, one needs to know or estimate the error covariance matrix, C_ϵ. Harville [1977] and Friston et al. [2002b] consider linear Gaussian models in which the error covariance can be specified in terms of hyperparameters, λ_i, where $C_\epsilon = \sum_i \lambda_i Q_i$. The matrices Q_i are known as 'covariance components' and specify the prior covariance structure, such as non-sphericity. These hyperparameters can be estimated using restricted maximum likelihood (ReML) or, equivalently, in an Expectation-Maximisation (EM) algorithm as follows [Friston et al., 2002b]:

$$
\begin{aligned}
&\textit{until convergence } \{ \\
&\textbf{E-step} \\
&\quad C_\epsilon = \textstyle\sum_i \lambda_i Q_i \\
&\quad P = (X^T C_\epsilon^{-1} X + \Sigma^{-1}) \\
&\quad \mu = C(X^T C_\epsilon^{-1} y + \Sigma^{-1} \eta) \\
&\textbf{M-step} \\
&\quad U = C_\epsilon^{-1} - C_\epsilon^{-1} X C X^T C_\epsilon^{-1} \\
&\quad g_i = -\tfrac{1}{2} tr(U Q_i) + \tfrac{1}{2} y^T U^T Q_i U y \\
&\quad H_{ij} = \tfrac{1}{2} tr(U Q_i U Q_j) \\
&\quad \lambda = \lambda + H^{-1} g \\
&\}.
\end{aligned}
\tag{2.6}
$$

The matrix U is the residual forming matrix, pre-multiplied by the error precision. This projector matrix 'restricts' the estimation of variance components to the null space of the design matrix. The quantities g and H are the first- and expected second-order derivatives (i.e. gradients and expected negative curvature) of the ReML objective function. This objective function is a special case of the variational free energy function described below (Section 2.3.2.1) [Friston et al., 2007a].

To summarise, the M-step calculates the hyperparameters by maximising the ReML objective function. In the E-step, the hyperparameter estimates are then used to update the posterior means and precisions. This procedure is repeated until convergence.

Once the moments are known (from Eq. 2.5 or Eq. 2.6), the posterior probability, p, that a particular effect or contrast c exceeds some threshold γ can be easily obtained:

$$
p = 1 - \Phi\left(\frac{\gamma - c^T \mu}{\sqrt{c^T C c}}\right), \tag{2.7}
$$

where Φ is the cumulative density function of the normal distribution. An image of these probabilities can be constructed by inverting a linear model in all voxels of the volume analysed. The corresponding maps are called posterior probability maps (PPMs) [Friston and Penny, 2003] and comprise an alternative approach to the t- and F-maps used in SPM.

2.2.3 Non-linear models

This section now describes the case of Bayesian model inversion for non-linear models. Whilst there are many nonlinear models and approaches for optimising them in the statistical literature [Gelman et al., 1995; Girolami, 2008; W. H. Press and Vetterling, 1992], this section focuses on the Bayesian optimisation algorithm used in this thesis.

A non-linear model, such as the biophysical models used in Chapters 5 and 6, can be linearised by expanding the observation equation about a working estimate μ_θ of the conditional mean. The prior density is assumed to be Gaussian with mean η and covariance Σ:

$$\begin{aligned} y &= g(\theta) + \epsilon \\ g(\theta) &\approx g(\mu_\theta) + J \cdot (\theta - \mu_\theta), \end{aligned} \tag{2.8}$$

such that $J = \frac{\partial g(\mu_\theta)}{\partial \theta}$, $y - g(\mu_\theta) \approx J \cdot (\theta - \mu_\theta) + \epsilon$ and $\epsilon \sim N(0, C_\epsilon)$. In the non-linear models used here, the error covariance is assumed isotropic over the predictions $C_\epsilon = \lambda I$, where I is the identity matrix.

The linearised model, Eq. 2.8, can be used in a Variational Laplace (VL) optimisation scheme. This assumes that the posterior is Gaussian and VL provides iterative updates of its moments: $q(\theta) \equiv q(\theta|y, m) = N(\mu, C)$. VL is a generic approach to estimate the posterior density, and can be formulated by analogy with statistical physics as a gradient ascent on the 'negative Free Energy', $F(m)$, of the system. The full derivation of the algorithm is described in Friston et al. [2007a]. Section 2.3.2.1 (see below) describes the free energy in more detail, with a focus on how this quantity can be used for model selection after model inversion.

The fixed-form Variational Laplace algorithm can also be regarded as an EM scheme [Friston et al., 2007a], of the form of Eq. 2.6, if one assumes that the variance hyperparameters have a point mass. The E-step performs an ascent on the variational free energy, F, to optimise the conditional moments of the local linear approximation (Eq. 2.8), whilst in the M-step, the hyperparameters λ are updated again by maximising F:

$$\begin{aligned} \text{E-step } \mu &= \arg\max_{\mu} F(\mu, \lambda) \\ \text{M-step } \lambda &= \arg\max_{\lambda} F(\mu, \lambda). \end{aligned} \tag{2.9}$$

Note that the posterior precision is not explicitly updated in the E-step because it is a function of the posterior mean. This type of EM scheme is based on earlier work by Neal and Hinton [1996];

The VL approximation to the posterior density over the model parameters has been verified using Markov Chain Monte Carlo (MCMC), in the context of DCM for fMRI Chumbley et al. [2007]. MCMC schemes are, in general, more computationally intensive but allow one to estimate the posterior density without assuming it has a particular form.

2.3 Bayesian model selection

2.3.1 Posterior model probabilities

After model inversion, Bayesian model selection proceeds by first assigning a prior probability to each model m_j, $p(m_j)$, in the comparison set, M. After observing data, y, BMS then uses the model evidence, or marginal likelihood, for each model, $p(y|m_j)$, and the prior to determine the posterior model probability, $p(m_j|y)$. This posterior probability is obtained through Bayes' rule, as follows:

$$p(m_j|y) = \frac{p(y|m_j)p(m_j)}{\sum_k^{n_M} p(y|m_k)p(m_k)}, \tag{2.10}$$

where the sum in the denominator is over all models that are being compared, and n_M is the total number of models. Model selection corresponds to choosing the model that most likely generated the data, i.e. the model m_j that maximises the posterior $p(m_j|y)$. If no model is favoured *a priori* then $p(m)$ is a uniform distribution, i.e. $p(m_j) = 1/n_M$, and the model with the highest posterior probability is also the model with the highest evidence, $p(y|m_j)$ [Kass and Raftery, 1995]. Therefore, the model evidence is often the fundamental object of interest in BMS.

The following sections describe the evidence in more detail, including different procedures for approximating this quantity.

2.3.2 Model evidence

The model evidence, $p(y|m_j)$, is the probability of obtaining observed data, y, given model m_j, belonging to model space M. As mentioned in the previous section, this quantity is at the heart of BMS, but, in general, it is not straightforward to compute, since this computation involves integrating out the dependency on the model parameters, θ:

$$p(y|m_j) = \int p(y|\theta, m_j)p(\theta|m_j)d\theta \tag{2.11}$$

Sampling or iterative analytic methods can be used to approximate the above integral. Here the focus is on how the model evidence can be computed for non-linear models. The evidence for linear models is then given as a special case.

A common approximation is the variational free energy approximation [Friston et al., 2007a; Penny et al., 2003; Sato et al., 2004; Stephan et al., 2009; Woolrich et al., 2004a]. In Chapter 5, the free energy is used to compare different neurovascular coupling models, whilst in Chapter 6, it is evaluated against another method for estimating and comparing dynamic causal models.

The following sections describe the free energy and other approximations to the model evidence.

2.3.2.1 Free energy approximation

Given model m_j with any parameters θ, using the rule of conditional probability and taking the log, the evidence can be written as:

$$\log p(y|m_j) = \log \frac{p(y, \theta|m_j)}{p(\theta|y, m_j)}. \tag{2.12}$$

If expectations are now taken with respect to the variational distribution, $q(\theta) \equiv q(\theta|y, m_j)$, the previous equation becomes:

$$\begin{aligned} \log p(y|m_j) &= \int q(\theta) \log \frac{p(y, \theta|m_j)}{p(\theta|y, m_j)} \frac{q(\theta)}{q(\theta)} d\theta \\ &= \int q(\theta) \log \frac{p(y, \theta|m_j)}{q(\theta)} d\theta + \int q(\theta) \log \frac{q(\theta)}{p(\theta|y, m_j)} d\theta, \end{aligned} \tag{2.13}$$

where the first term in the previous equation is known as the negative free energy

of the system, F_j, and the second term is the Kullback-Leibler (KL) divergence between the approximate posterior density, $q(\theta)$, and the true posterior, $p(\theta|y, m_j)$. Given these definitions, the log model evidence can be written as:

$$\log p(y|m_j) = F_j + KL(q(\theta)||p(\theta|y, m_j)). \tag{2.14}$$

Since the KL divergence between the approximate and true posterior is always positive, or zero when the densities are identical [Bishop, 2006], the log-evidence, $\log p(y|m_j)$, is therefore bounded below by F_j.

Equation 2.14 describes the fundamental relationship between evidence, free energy and KL divergence, which is used in the Variational Bayes (VB) optimisation framework [Beal and Ghahramani, 2003]. In VB, the moments of the posterior density $q(\theta)$ are updated by implicitly minimising the KL divergence. Through this iterative procedure F_j becomes an increasingly tighter lower bound on the log-model evidence.

In general, the negative free energy can be further decomposed into the sum of accuracy and complexity terms:

$$F_j = Accuracy(m_j) - Complexity(m_j). \tag{2.15}$$

The accuracy is defined as the average log-likelihood, whilst the complexity term is the KL divergence between the approximate posterior $q(\theta)$ and the prior density over the model parameters, $p(\theta|m_j)$:

$$\begin{aligned} Accuracy(m_j) &= \int q(\theta) \log p(y|\theta, m_j) d\theta \\ Complexity(m_j) &= KL(q(\theta)||p(\theta|m_j)). \end{aligned} \tag{2.16}$$

The complexity term therefore tends to increase with the number of model parameters, penalising complex models, but only if these extra parameters diverge from their priors [Penny, 2012].

Finally, under Gaussian assumptions (Laplace approximation), $q(\theta) = N(\mu, C)$ and $p(\theta|m_j) = N(\eta, \Sigma)$, the variational free energy can be written as:

$$\begin{aligned} F_j &= -\frac{1}{2}e_y^T C_\epsilon^{-1} e_y - \frac{1}{2}\log|C_\epsilon| - \frac{N_y}{2}\log 2\pi \\ &- \frac{1}{2}e_\theta^T \Sigma^{-1} e_\theta - \frac{1}{2}\log|\Sigma| + \frac{1}{2}\log|C|, \end{aligned} \tag{2.17}$$

where $e_y = y - g(\mu)$, $e_\theta = \mu - \eta$ and N_y is the number of observations.

In the case of general linear models, $g(\theta) = X\theta$, the free energy becomes equal to the ReML objective function. If the model contains unknown hyperparameters λ, with prior $p(\lambda|m) = N(\eta_\lambda, \Sigma_\lambda)$ and posterior $q(\lambda) = N(\mu_\lambda, C_\lambda)$, three more terms can be added to F_j, in order to account for these parameters:

$$\begin{aligned} F_j &= -\frac{1}{2}e_y^T C_\epsilon^{-1} e_y - \frac{1}{2}\log|C_\epsilon| - \frac{N_y}{2}\log 2\pi \\ &- \frac{1}{2}e_\theta^T \Sigma^{-1} e_\theta - \frac{1}{2}\log|\Sigma| + \frac{1}{2}\log|C| \\ &- \frac{1}{2}e_\lambda^T \Sigma_\lambda^{-1} e_\lambda - \frac{1}{2}\log|\Sigma_\lambda| + \frac{1}{2}\log|C_\lambda|, \end{aligned} \tag{2.18}$$

where $e_\lambda = \mu_\lambda - \eta_\lambda$. Model selection can then proceed using the free energy F_j as a surrogate for the log-evidence. We note that, as shown in the Appendix C of Wipf and Nagarajan [2009], the Laplace approximation, Eq. 2.18, is not equal to the Variational Free Energy in Eq. 2.14 (it contains extra terms that can be either positive or negative, see Wipf and Nagarajan [2009] for the derivation) and therefore the lower bound property might no longer hold under this approximation. That is, the Laplace approximation is not strictly a lower bound on the log-model evidence. Nevertheless, it provides a very informative and useful approximation to the free energy, which can still be used as a model selection criterion [Friston et al., 2007a; Penny, 2012].

2.3.2.2 Other approximations

Other approximations to the model evidence exist, including the computationally more expensive Annealed Importance Sampling (AIS) method [Beal and Ghahramani, 2003], and the simpler but potentially less accurate Bayesian Information Criterion (BIC) [Schwarz, 1978] and Akaike Information Criterion (AIC) [Akaike, 1973] measures, respectively:

$$\begin{aligned} BIC &= Accuracy(m_j) - \frac{n_p}{2}\log N_y \\ AIC &= Accuracy(m_j) - n_p, \end{aligned} \tag{2.19}$$

where n_p is the number of parameters θ in the model. In extensive simulations of graphical model structures, Beal and Ghahramani [2003] found that the variational approach outperformed BIC and AIC, at relatively little extra computational cost, and approached the performance of AIS, but with much less computational cost. In addition, Penny [2012] shows that for the case of comparing DCMs, the free energy approach also performs better than either AIC or BIC. Kass and Raftery [1995] show that AIC and BIC are biased to complex and simple models, respectively .

More recently, Friston and Penny [2011] have proposed a post-hoc approximation to the evidence, which is computed by optimising only the very largest of a set of models. This approach can be viewed as a generalisation of the well-known Savage-Dickey ratios [Dickey, 1971]. Chapter 6 derives the post-hoc model evidence and compares this approach to the free energy approximation described above, using dynamic causal models for fMRI data.

2.3.3 Bayes factors

Two models, m_i and m_j, can be compared using Bayes Factors, which are defined as the ratio of the corresponding model evidences. Equivalently, log-Bayes Factors are given by differences in log-evidences:

$$\log B_{ij} = \log p(y|m_i) - \log p(y|m_j). \tag{2.20}$$

The Bayes factor is a summary of the evidence provided by the data in favour of one scientific hypothesis, represented by a statistical model, as opposed to another. Bayes factors have been stratified into different ranges deemed to correspond to different strengths of evidence [Kass and Raftery, 1995]. 'Strong' evidence, for example, corresponds to a BF of over 20 (log-BF over 3) in favour of model m_i when compared to model m_j. Bayes factors can also be directly interpreted as odds ratios where B_{ij} = 100, for example, corresponds to odds of 100 to 1.

Bayes factors can be used to convert a prior odds ratio into a posterior odds ratio:

$$\frac{p(m_i|y)}{p(m_j|y)} = B_{ij}\frac{p(m_i)}{p(m_j)} \tag{2.21}$$

For equal prior odds, as is the case for uniform priors, the posterior odds is equal to the Bayes factor. From this, the equivalent posterior probability of hypothesis m_i can be computed as follows:

$$p(m_i|y) = \frac{1}{1+\frac{1}{B_{ij}}}, \tag{2.22}$$

and a posterior model probability greater than 0.95 is equivalent to a Bayes Factor greater than 20.

Importantly, unlike classical inference and p-values, Bayes factors can be used to compare non-nested as well as nested models. They also allow one to quantify evidence in favour of a null hypothesis.

So far, the theory described applies to single subject analyses. The following sections describe Bayesian model selection at the group level.

2.3.4 Group studies

Neuroimaging studies usually require data from a group of subjects. This is because the signals are of small magnitude and may not be detectable in every subject. Collecting data for a group of subjects therefore allows one to assess group effects in relation to the variability over the group.

2.3.4.1 Fixed-effects

Until very recently, most group studies have adopted what is known as the Group Bayes Factor (GBF) approach [Stephan and Penny, 2007]. After inverting all models for all subjects, the GBF can be obtained by simply multiplying the individual BFs for all N subjects (assuming subjects are independent):

$$\begin{aligned} GB_{ij} &= \prod_{n=1}^{N} B_{ij}^n \\ \log GB_{ij} &= \sum_{n=1}^{N} \log p(y_n|m_i) - \sum_{n=1}^{N} \log p(y_n|m_j), \end{aligned} \tag{2.23}$$

where the subscripts i and j denote the i-th and j-th models being compared, and y_n is the data from subject n. The log GBF is therefore simply the difference of the log-model evidences aggregated over subjects. Although this is a straightforward method for model selection and has been used in a number of neuroimaging studies [Stephan et al., 2007b; Summerfield and Koechlin, 2008], Stephan et al. [2009] have recently shown that the Group Bayes Factor approach corresponds to what is more generally known as a Fixed Effects (FFX) analysis. The FFX approach can be understood as a special case of a more general Random Effects (RFX) model inference approach, as described below.

2.3.4.2 Random-effects

Random effects model inference [Stephan et al., 2009] is based on model frequencies, $r = [r_1, ..., r_{n_M}]$, where $0 \leq r_k \leq 1$ and $\sum_{k=1}^{n_M} r_k = 1$, which represent frequencies of models used in the population at large (Figure 2.2). Stephan et al. [2009] have developed a hierarchical model for making inferences on the posterior density of the model frequencies, $p(r|Y)$, given the data from all subjects, Y. This method can be viewed as a RFX approach, in which a (potentially different) model is assigned to each member of the group (Figure 2.2B). In other words, the assignment of models to subjects is treated as a random process. In FFX, the same model is assigned to all subjects (Figure 2.2A).

The corresponding random variables are drawn from a density, $p(r|\alpha)$, which then defines a distribution on how likely it is that model k generated the data for subject n, $p(m_{nk} = 1) = r_k$, where $m_{nk} \in \{0, 1\}$ and $\sum_{k=1}^{n_M} m_{nk} = 1$. Because, for each subject, this latter distribution has a multinomial form (i.e. each subject uses either model $k = 1, 2, ...n_M$) it is natural to choose $p(r|\alpha)$ as a Dirichlet density, as the Dirichlet is conjugate to the multinomial [Bernardo and Smith, 2001]. The parameters of this Dirichlet, $\alpha = [\alpha_1, ..., \alpha_{n_M}]$, are related to the unobserved 'occurrences' of the models in the population. FFX is then a special case of RFX, when $r_k = 1$ for the model with the highest summed evidence over the group and $r_k = 0$ otherwise. These probabilities indicate that all subjects use the same model, whereas for RFX subjects can use different models.

The same authors then describe an estimation procedure to invert this hierarchical model and estimate the posterior distribution over r. Briefly, this optimisation scheme begins by assuming that each model has been 'observed' once, $\alpha_0 = [1, ..., 1]$, and

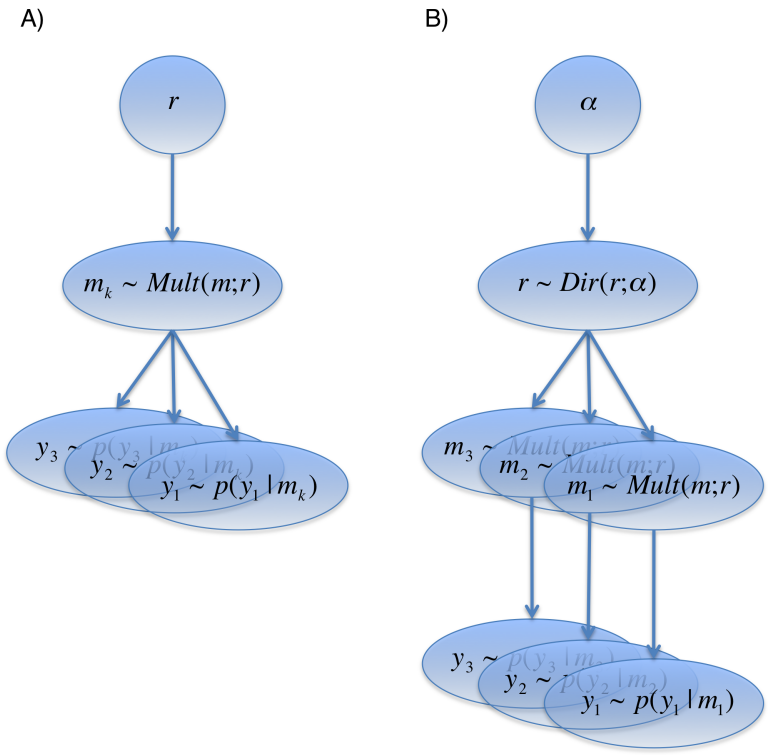

Figure 2.2: Graphical models underlying (A) Fixed and (B) Random Effects inference on model space at the group level. FFX assigns a model, drawn using r, to be used by all members of the group, whilst for RFX a (potentially different) model is assigned to each member of the group. $Mult(m;r)$ and $Dir(r;\alpha)$ refer to multinomial and Dirichlet densities. See the main text for a detailed explanation of the two different inference approaches.

proceeds by updating estimates of α until convergence. The following pseudo-code schematises this iterative procedure and the quantities computed at each step:

$$\begin{aligned}
&\alpha = \alpha_0 \\
&\text{until convergence} \\
&\qquad \text{compute } g_{nk} \text{ (see Eq. 2.25)} \\
&\qquad \text{compute } \beta \text{ (see Eq. 2.26)} \\
&\qquad \text{update } \alpha = \alpha_0 + \beta \\
&\text{end.}
\end{aligned} \tag{2.24}$$

In the first step the normalised posterior belief that model k generated the data from subject n, g_{nk}, is computed using the following equations:

$$\begin{aligned}
u_{nk} &= \exp(\log p(y_n|m_{nk}) + \Psi(\alpha_k) - \Psi(\alpha_S)) \\
u_n &= \sum_{k=1}^{n_M} u_{nk} \\
g_{nk} &= \frac{u_{nk}}{u_n},
\end{aligned} \tag{2.25}$$

where $\log p(y_n|m_{nk})$ is the log-model evidence from subject n and model k, Ψ is the digamma function, $\Psi(\alpha_k) = \partial \log \Gamma(\alpha_k)/\partial \alpha_k$, and $\alpha_S = \sum_k \alpha_k$. In the next step, the expected number of subjects whose data are believed to have been generated by model k is computed for all models:

$$\beta_k = \sum_n g_{nk}. \tag{2.26}$$

Finally, using the result from the previous step, the α parameters are updated (Eq. 2.24).

After optimisation, the posterior distribution $p(r|Y;\alpha)$ can be used for model inference at the group level. One can, for instance, use this distribution to compute the expected model frequencies, $\langle r_k \rangle$:

$$\langle r_k \rangle = \alpha_k/(\alpha_1 + ... + \alpha_{n_M}), \tag{2.27}$$

Another option is to use $p(r|Y;\alpha)$ to compute an exceedance probability, φ_k, which corresponds to the belief that model k is more likely than any other (of the

n_M models compared), given the data from all subjects:

$$\varphi_k = p(\wedge_{j \neq k} r_k > r_j | Y; \alpha). \tag{2.28}$$

Exceedance probabilities are particularly intuitive when comparing just two models (see for example Figure 4.5B) as they can be written:

$$\varphi_1 = p(r_1 > r_2 | Y; \alpha) = p(r_1 > 0.5 | Y; \alpha). \tag{2.29}$$

Finally, it is important to note that if there are reasons to believe that the optimal model is identical across subjects, then the FFX approach is entirely valid. This assumption is justified when studying, for instance, a basic physiological mechanism that is unlikely to vary across subjects, such as the role of forward and backward connections in visual processing.

2.4 Conclusion

This chapter has reviewed the model inversion and selection methods used in the following Results chapters. Model selection, and in particular Bayesian model selection, plays a key role in any scientific discipline by providing a principled way of choosing between competing hypothesis for explaining the observed data. The coming years will probably see the development of more efficient ways, such as the use of stochastic algorithms, of searching through very large numbers of models. These developments will have particular impact on connectivity research, where they will facilitate network discovery in very large model spaces.

Chapter 3

Estimating the transfer function from neuronal activity to BOLD using simultaneous EEG-fMRI

3.1 Introduction

Functional magnetic resonance imaging (fMRI), with blood oxygenation level dependent (BOLD) contrast, is an established method for making inferences about regionally specific activations in the brain [Frackowiak et al., 2003]. However, the relationship between BOLD and neuronal activity is still under debate, in particular, it is still unclear how the hemodynamic response is influenced by the temporal dynamics of the underlying neuronal activity.

One of the approaches used to study this relationship is to combine information from hemodynamic measures such as fMRI and electrophysiological measures, such as electroencephalography (EEG) and magnetoencephalography (MEG) (see Chapter 1). EEG and MEG are well-established non-invasive techniques, and are well suited to studying the temporal dynamics of neuronal activity, since they provide direct measurement of this activity with high temporal resolution [Hämäläinen et al., 1993].

In humans, the study of correlations between EEG and fMRI signals has been pioneered by epilepsy researchers, such as Lemieux et al. [2001] and Salek-Haddadi et al. [2002]. However, most of our present knowledge about neurovascular coupling comes from animal research, as described in Chapter 1. This research supports the emerging consensus in which BOLD is thought to result from pre-synaptic activity

and the release of neurotransmitters, in particular glutamate [Friston, 2008]. This release triggers a response in surrounding glial cells, especially astrocytes, leading to the generation of vasodilatory signals and consequently BOLD [Nair, 2005]. As well as indirectly causing BOLD, glutamate will increase post-synaptic activity and therefore the local field potentials (LFP). Increases in LFP frequency would therefore be accompanied by faster glutamate recycling and consequently a larger BOLD signal.

Whilst the above physiological perspective would suggest that BOLD should be sensitive to the frequency content of neuronal activity, results from the neuroimaging literature are not completely clear cut. For example, some studies (see next paragraph) suggest that BOLD is mainly dependent on the total energy, or total spectral power, of neuronal activity. Others (see next but one paragraph), suggest that BOLD is sensitive to a certain range of frequencies or results from more complicated dynamics.

Among those proposing a relationship between BOLD and total neuronal power, Wan et al. [2006] have found significant correlation between the mean power (mean square current source density estimates during visual stimulation) of source reconstructed EEG data in human primary visual cortex and a neuronal efficacy parameter, derived from fitting a Balloon model to fMRI data (see Chapter 1). Similarly, Nangini et al. [2008] propose that the energy density, as measured by the square of the equivalent current dipole (ECD) waveforms from source reconstructed MEG data, is a better representation for the neuronal input functions, than the stimulus functions conventionally used in convolution models for the analysis of fMRI data [Friston et al., 1995]. In addition to these studies, theoretical models for integrating EEG/MEG and fMRI (Babajani and Soltanian-Zadeh [2006]; Nunez and Silberstein [2000]; Trujillo-Barreto et al. [2001]) assume a relationship between indices of neuronal activity and BOLD that is independent of the frequency of this activity. For instance, Babajani and Soltanian-Zadeh [2006] use a neural mass model of neuronal activity and propose that the squared post-synaptic membrane potential from both excitatory and inhibitory cells from a given cortical area drives increases in cerebral blood flow, and consequently BOLD.

Among those proposing a relation between BOLD and the frequency structure of electrophysiological signals Goldman et al. [2002], Moosmann et al. [2003] and Laufs et al. [2003] have shown that reductions in ongoing scalp EEG alpha power in humans correlate with increases in BOLD activity. Lachaux et al. [2007] have found, using intra-cranial recordings in epileptic patients, a close spatial correspondence between

regions of fMRI activation and sites showing EEG energy variation in the gamma band. Mukamel et al. [2005] have found positive correlations between LFPs and BOLD at high gamma-range frequencies ([40, 130] Hz) and negative correlations at low/alpha-range frequencies ([5, 15] Hz) in auditory cortex of neurosurgical patients. In addition, Niessing et al. [2005] have shown that fluctuations in hemodynamic response tightly correlate with the power of LFP oscillations, recorded in cat primary visual cortex, in the same high frequency (gamma) range.

Kilner et al. [2005] note that from the perspective of fMRI, neuronal activation is proportional to relative metabolic demands, or rate of energy dissipation (1/s units). In terms of EEG, the effect of activation is to shift the spectral profile toward higher frequencies (1/s units) with a reduction in amplitude. This led Kilner et al. [2005] to propose a 'Heuristic' model that links these two observations via a dimensionality analysis. This Heuristic specifies that BOLD activations are accompanied by an increase in the 'average' frequency of EEG neuronal activity, where average is defined in the root mean square (RMS) sense. Thus increases in higher frequencies, such as the gamma range, relative to lower frequencies, such as the alpha range, would lead to increases in BOLD. Conversely, increases in alpha relative to gamma would lead to decreases in BOLD. The equations of the 'Heuristic' model and its assumptions are summarised in the Appendix.

Moreover, using data from Niessing et al. [2005] the Heuristic model has been shown to provide a better fit than a model based on gamma correlation alone [Kilner et al., 2007]. In similar spirit to the idea underlying the Heuristic, Laufs et al. [2006a] have found that BOLD deactivations in humans are associated with increases in the ratio between theta and alpha bands (measured with scalp EEG), and that these deactivations cease when there is a decrease in this ratio and an increase in the beta/alpha ratio.

More recently, Goense and Logothetis [2008] used simultaneous intra-cortical LFP-BOLD recordings and a multiple regression model in which activity in many different frequency bands, covering the entire LFP range of frequencies, were used to predict BOLD activity in alert behaving monkeys. The results showed that all bands explained a significant part of the BOLD response.

The link between neuronal activity and BOLD has been investigated at both a microscopic scale, using invasive, co-localised recordings in animals (e. g. Goense and Logothetis [2008]; Logothetis et al. [2001]; Niessing et al. [2005]), and at a macroscopic scale using simultaneous EEG-fMRI in humans [Goldman et al., 2002; Laufs et al.,

2003; Lemieux et al., 2001; Moosmann et al., 2003]. A problem with the macroscopic approach is that the electrophysiological measure, EEG, is not co-localised with BOLD. This issue can be addressed by the use of principal component analysis (PCA) [Laufs et al., 2006b], independent component analysis [Eichele et al., 2005, 2009] or source reconstruction [Wan et al., 2006]. A summary of the above findings is provided in Table 1.2 of Chapter 1.

In this chapter we use simultaneous EEG-fMRI in humans and employ a visual flicker stimulation paradigm to elicit evoked activity in sensory cortex. As scalp EEG measures the activity of multiple distributed neuronal processes we used a PCA approach to isolate activity that was primarily related to the stimulus paradigm. The resulting time series was then used as a surrogate for neuronal activity.

We then regressed the fMRI data onto convolved features of the power spectrum of the first principal component of the EEG data. We use a standard statistical parametric mapping (SPM) approach employing F-tests to compare models embodying different transfer functions. These are (i) a total power model (ii) a frequency response model, comprising multiple regression onto power in different frequency bands, and (iii) a Heuristic model in which BOLD is predicted by the RMS EEG frequency.

This chapter is structured as follows. In Section 3.2 we describe the experimental paradigm and the simultaneous acquisition of EEG and fMRI data. We also describe the preprocessing steps used for artefact removal, define the transfer functions investigated and the methods used to compare models. Section 3.3 presents the results from the SPM analysis and in Section 3.4 these results are discussed in light of previous results from the literature.

3.2 Materials and methods

3.2.1 Subjects and task

Three healthy volunteers (three male, mean age 35 ± 4 years) participated in the study after giving informed consent. Subjects were exposed to visual flicker stimuli of a number of different frequencies. A reversing black and white checkerboard (11 x 11 squares, size 13 cm x 13 cm) was delivered via a computer monitor (60 Hz refresh rate) and projected on a screen positioned 47 ± 1 cm from a 45° mirror located 11 ± 3 cm from the subject (visual angle = 6.5 ± 0.5°). The stimulation frequencies used were 2.00, 3.75, 5.00, 6.00, 7.50, 10.00, 15.00 and 30.00 Hz (screen

reversing frequencies correspond to 2 times the value of the stimulation frequencies; in Chapter 5 we use reversing frequencies). Stimuli were delivered in epochs of 5 scans (15.3 s), followed by periods of 15.3 s of rest (blank screen), and the order of stimulus blocks (eg. 10 Hz, 6 Hz, 5 Hz etc.) was randomised. Subjects were instructed to view a fixation cross which was visible during both rest and stimulus periods, and no overt response was required in either condition. Three consecutive sessions of the same experimental task were recorded for each subject. Although luminance levels were not held constant for the different flicker frequencies, the variations in luminance were measured using a lux meter placed in front of the visual display unit. This allowed luminance variations to be regressed out during subsequent statistical analyses, when required.

As the aim of our experiment was to investigate a basic physiological mechanism, the neurovascular coupling, the paradigm used here was designed to induce a large electrophysiological response in sensory cortex. For this reason, inter-subject variability was found to be low, i.e. the response to the stimulation paradigm, as observed in fMRI and EEG signals (see below), was found to vary little from subject to subject. It was therefore appropriate [Penny and Holmes, 2006] to acquire data from a small number of subjects (three) and to report results in the form of case studies, as well as to summarize these results using fixed effects SPMs (see below). This follows the precedent of Wan et al. [2006] who also used a case study approach with a small number of subjects (five).

3.2.2 EEG acquisition

EEG was acquired simultaneously with fMRI using a synchronized imaging protocol [Mandelkow et al., 2006] and an MR-compatible BrainAmp amplifier and BrainCap EEG cap with ring Ag/AgCl electrodes (Brainproducts GmbH, Munich, Germany). Raw EEG was sampled at 5 kHz and a low pass filter (cut off frequency: 1 kHz) was used. This system provided 29 EEG channels, 2 EOG channels, and 1 ECG channel. The electrodes were distributed according to the 10/20 system, and the reference electrode was located between Fz and Cz. EEG was also recorded outside of the MRI environment (in a dark and acoustically isolated room), so that the effect of MRI-induced artefacts and their removal could be assessed. We additionally measured the pulse using a pulse oxymeter attached to the subject's finger and the locations of the EEG electrodes were digitised with a Polhemus digitiser.

3.2.3 fMRI acquisition

Images were acquired from a 1.5 T whole-body scanner (Magnetom Sonata, Siemens Medical, Erlangen, Germany) operated with its standard body transmit and CP head receive coil. The manufacturers standard automatic 3D-shim procedure was performed at the beginning of each experiment. The scanner produced T2*-weighted images with a single-shot gradient-echo EPI sequence. Whole brain images consisting of 34 contiguous transverse slices, on a 64-by-64 grid, were acquired every 3.06 seconds resulting in a total of 320 functional scans for each of the three sessions of each subject (slice thickness = 2 mm, gap between slices = 1 mm, repetition time TR = 90ms, flip angle = 90°, echo time TE = 50 ms, field of view FOV = 192 × 192 mm^2, and therefore 3 × 3 × 3 mm voxel resolution). Whole-brain structural scans were also acquired using a T1-weighted 3D-Modified Driven Equilibrium Fourier Transform (MDEFT) sequence [Deichmann et al., 2004] in 176 sagittal partitions with an image matrix of 256 × 256 (TR = 12 ms, TE = 4 ms, flip angle = 23°, and voxel size 1 × 1 × 1 mm).

3.2.4 EEG data analysis

Acquisition of EEG in the MRI environment induces Gradient and Cardiac related artefacts, such as the Ballistocardiogram artefact [Goldman et al., 2000], as discussed in Chapter 1. The data acquired inside the scanner were corrected off-line using facilities in the Brain Vision Analyzer software package (Brainproducts GmbH, Munich, Germany) [Allen et al., 2000]. First, the Gradient artefact was removed via mean subtraction with template drift compensation. Cardiac related artefacts were then removed by subtracting the first three principal components that were time-locked to pulse oxymeter readings. EEG data acquired outside the scanner were not processed in this way. Both the data acquired inside and outside the scanner were then high-pass filtered (0.5 Hz) to reduce slow drifts in the signal.

After MR-related artefact removal and filtering, the inside and outside EEG data were visually inspected for other artefacts, such as eye-blinks, as well as movement related artefacts. Due to their proximity to the subjects' eyes, the Fp1 and Fp2 electrodes contained too many eye-blink artefacts to be included in the analysis.

After visual inspection, the EEG data from the remaining channels were then processed to form a single representative 'scalp EEG' time series, by projecting the data onto a subspace defined by its first principal eigenvector u_1.

In previous work Moosmann et al. [2003] and Laufs et al. [2003], have generated a single representative time series by computing the mean over a subset of activated electrodes (e.g. 01, 02, P1, P2). We have used a spatial eigendecomposition method because this data driven approach produces the single time series which, out of all possible linear projections, captures most variance in the original data. However, as brain activity in our paradigm is primarily driven by activity in visual cortex this spatial eigenmode is primarily loaded onto posterior electrodes, as is shown below.

The principal eigenvectors can be computed from a singular value decomposition (SVD) of the data. If Y is an $n_e \times n_t$ matrix of EEG data, with n_e electrodes and n_t time points, then an SVD gives $Y = USV^T$, and the projection is given by: $\tilde{y} = u_1^T Y$, where u_1 is the first column of U.

To investigate the spectral properties of the scalp signal, $\tilde{y}(t)$, we decomposed it into the time-frequency domain. This decomposition was obtained by convolving the signal with Morlet wavelets, G, where for each time point t and frequency f:

$$G(f,t) = A \exp(-t^2/(2\sigma_t^2)) \exp(2i\pi ft), \tag{3.1}$$

where $A = (\sigma_t\sqrt{\pi})^{-1/2}$, $\sigma_t = 1/(2\pi\sigma_f)$, $\sigma_f = f/R$, and $R = 7$ is the "wavelet factor". The time-varying power of the signal around frequency f, is then given by the squared modulus of the convolution [Tallon-Baudry and Bertrand, 1999]:

$$P(f,t) = |G(f,t) * \tilde{y}(t)|^2, \tag{3.2}$$

and the power spectrum for all frequencies and time points can be represented by the matrix P with dimensions $n_f \times n_t$, where n_f is the number of frequencies.

3.2.5 Transfer functions

From the spectrum of the EEG data, P, we constructed regressors defining the different transfer functions we were interested in comparing. These represent the functional link between neuronal activity and BOLD.

The first model, motivated by the result of Wan et al. [2006], assumes that neurovascular coupling is a power transducer. To this end we derived a feature corresponding to the 'Total Power' in the scalp EEG time-series. This was obtained by summing the EEG power over all frequencies analyzed ([1, 40] Hz):

$$q_{TP}(t) = \sum_{f=1}^{n_f} P(f,t). \tag{3.3}$$

The second model, following Goense and Logothetis [2008], assumes that BOLD is best explained by a linear combination of activity in different frequency bands. We refer to this as the 'Frequency Response' model and consider three variants, each with a different number of frequency bands. These comprise (i) three bands of low frequencies [1, 7] Hz, alpha frequencies [8, 15] Hz and higher frequencies [15, 40] Hz, (ii) five bands of delta [1, 4] Hz, theta [4, 8] Hz, alpha [8, 13] Hz, beta [13, 30] Hz and lower gamma [30, 40] Hz activity and (iii) eight bands of 5 Hz each, from 1 to 40 Hz. The time-series for each band were obtained by summing the power in the corresponding frequency interval, $b = [f_{min},\ f_{max}]$:

$$q_{FR}(t)_b = \sum_{f=f_{min}}^{f_{max}} P(f,t). \tag{3.4}$$

The resulting time-series for each band, b, correspond to different columns of the same design matrix (see below).

The third model, which we refer to as the 'Heuristic' model based on Kilner et al. [2005], assumes that BOLD is best explained by a linear convolution of the 'Root Mean Squared Frequency' (RMSF) function. This is given by

$$q_{RMSF}(t) = \sqrt{\sum_{f=1}^{n_f} f^2 \tilde{P}(f,t)}, \tag{3.5}$$

where $\tilde{P}$ is the corresponding normalised power spectrum of the representative scalp time series (at each time point, t, power in each frequency is divided by the total power over all frequencies, i.e. at each time point the normalised power spectrum sums to 1 over all frequencies). This function describes how changes in the relative power of the different frequencies in the EEG spectrum could be associated with changes in BOLD.

We also investigated two variants of the Heuristic. The first, uses the un-normalised power spectrum P, instead of $\tilde{P}$:

$$q_{uRMSF}(t) = \sqrt{\sum_{f=1}^{n_f} f^2 P(f,t)}. \tag{3.6}$$

We refer to this as the 'un-normalised Heuristic' (u-Heuristic). Second, to test for the importance of the non-linearity introduced by the square root in the RMSF function we defined the function,

$$q_{MSF}(t) = \sum_{f=1}^{n_f} f^2 \tilde{P}(f,t), \tag{3.7}$$

which is a linear version of Eq. 3.5. We refer to this as the 'linear Heuristic' (l-Heuristic) model.

To further test the importance of the non-linearity we defined another function based on a linear convolution of the 'Mean Frequency' (MF) of the EEG signal:

$$q_{MF}(t) = \sum_{f=1}^{n_f} f \tilde{P}(f,t). \tag{3.8}$$

Finally, we constructed one last frequency-independent transfer function purely based on variations of amplitude in the EEG signal, as captured by the global field power (GFP). The GFP corresponds to the root-mean-square deviations between all electrodes in a given potential field [Skrandies, 1995]:

$$q_{GFP}(t) = \sqrt{\sum_{i=1}^{n_e} (U_i(t) - \bar{U}(t))^2}, \tag{3.9}$$

where $\bar{U}(t) = \frac{1}{n_e} \sum_{j=1}^{n_e} U_j(t)$ is the mean of the potential across electrodes at a given time point. This is a reference-free measure and allowed us to compare the previously described transfer functions, which are all based on the power spectrum of the EEG data, with a measure based simply on the amplitude of the EEG signal. This is also the only function constructed using data from all electrodes, instead of the virtual electrode obtained using PCA.

For each of the above models, the time series were convolved with an informed basis set to accommodate variability in the hemodynamic response. This basis set includes the canonical hemodynamic response function (HRF), as well as its first temporal and dispersion derivatives (see Chapter 1). The two derivative regressors

allow for variations, across subjects and across the brain, in the peak response time and duration of the hemodynamic response. The temporal derivative, for example, allows for peak responses that are approximately one second earlier or later than is usual. The convolved time series were then downsampled to match the fMRI sampling rate, and served as regressors of interest in the subsequent general linear model (GLM).

As we are using an informed basis set with 3 basis functions, each of the Total Power, Heuristic, u-Heuristic, l-Heuristic, MF and GFP models are implemented using 3 design matrix columns. There are therefore 3 corresponding regression coefficients of interest to estimate for each of these models. The Frequency Response model is implemented with 9, 15 or 24 columns for the 3, 5 or 8 band-model, respectively. The coefficients of interest, as well as the total number of parameters estimated for each function are summarised in Table 3.1. Whilst some of these models yield a large number of parameters, this is readily accommodated because there are 9 sessions of scanning with 100 data points each, yielding a total of 900 data points (see below). However, we note that although we have 900 data points, these are not all independent due to the temporal correlations in the BOLD time series. In the GLM framework, temporal correlation is accounted for using an autoregressive AR(1) model during classical ReML parameter estimation [Friston et al., 2007b].

3.2.6 fMRI data analysis

The fMRI data were pre-processed with SPM8 software (http://www.fil.ion.ucl.ac.uk/spm/) implemented in Matlab (The Mathworks, Inc.). The first five scans of each session were discarded, and the pre-processing steps included: (a) realigning the images to the first scan and coregistering the structural scan of each subject with the mean functional image from all sessions; (b) correcting for differences in acquisition time between slices (known as slice-timing correction) and normalising all the functional and structural scans to a standard EPI template based on the Montreal Neurological Institute (MNI) reference brain in Talairach space [Talairach and Tournoux, 1988] (c) smoothing the functional images (Gaussian kernel, 8 mm half width). The movement parameters[1] obtained from the realignment step were included in the subsequent GLM analyses as confounding covariates (Table 3.1). The data were also high-pass

[1]The movement parameters correspond to the three translations and rotations that are applied to each scan in order to realign it with a reference image.

Table 3.1: Summary of model comparisons and corresponding number of estimated parameters. For one session: n_R is the number of regressors of interest for each transfer function; n_{BF} is the number of basis functions, which is always 3 (canonical HRF, temporal and dispersion derivative); n_C is always 7 and corresponds to the number of confounds (6 motion parameters and 1 mean regressor); n_P is the total number of parameters to be estimated for each comparison.

	Model comparisons	$n_R \times n_{BF} + n_C = n_P$	n_P (9 sess.)
i.	Heuristic	$1 \times 3 + 7 = 10$	90
	Total Power (TP)	$1 \times 3 + 7 = 10$	90
	Freq. Response 3 bands (FR_3)	$3 \times 3 + 7 = 16$	144
ii.	TP vs Heuristic	$(1 + 1) \times 3 + 7 = 13$	117
	TP vs FR_3	$(1 + 3) \times 3 + 7 = 19$	171
iii.	TP vs FR_3 vs Heuristic	$(1 + 3 + 1) \times 3 + 7 = 22$	198
iv.	TP vs FR_3 vs Heuristic vs GFP	$(1 + 3 + 1 + 1) \times 3 + 7 = 25$	225
v.	Heuristic vs FR_1	$(1 + 1) \times 3 + 7 = 13$	117
vi.	Heuristic vs u/l-Heuristic/MF	$(1 + 1) \times 3 + 7 = 13$	117
	Heuristic vs FR_5/FR_8	$(1 + 5/8) \times 3 + 7 = 25/34$	225/306

filtered, with a cut off period of 128 s.

We report analyses based on the first 100 scans of each session due to suspected movement-related (i.e. high amplitude and high-frequency) artefacts present in the EEG signal, after approximately 5 minutes of recording, in more than one session and subject. However, we later visually re-inspected the EEG signal and decided to include some of the previously discarded scans, and re-analysed the data using 200 scans per session. This new analysis yielded very similar results and strengthened the findings obtained with less data (see below).

For each subject we first looked at the effect of the experimental task. We used the onsets of the stimuli as regressors, and inferences based on the statistical parametric maps (SPMs) from a fixed effects group analysis were considered significant at $p < 0.05$, corrected for multiple comparisons using random field theory [Friston et al., 1995] (see Chapter 1). This preliminary SPM analysis, which used convolved visual stimulus onsets as regressors, is therefore a separate analysis than the one in which EEG-derived features were used as regressors. Inference was based on F-tests, which test for the additional variance explained by a set of regressors of interest (see Chapter 2). We also used these maps to generate a mask image, which we refer to as the 'BOLD activation mask'. This mask allows us to look at correlations between model predictors and BOLD, limited to the voxels activated by the checkerboard stimuli.

3.2.7 Model comparisons

In this section we describe the comparisons between transfer functions that were performed in order to investigate the link between neuronal activity and BOLD.

We began by looking at correlations between individual functions and the BOLD signal, by using these functions in *separate design matrices*. This was followed by a more formal comparison, which included regressors from multiple models in the *same design matrix*.

Inference in both cases was based on F-tests (see Chapter 2). In the first case we test for the effect of each model alone, i.e. without taking into account the rest of the models. This is to reproduce previously published results, in which each individual feature of neuronal activity has been linked to BOLD. In the second case, a significant F-statistic for a particular transfer function suggests that model explains BOLD variability that can't be explained by any of the other coupling functions in that design matrix [Friston et al., 1995]. This allows us to infer that one model is

better than another.

These tests were performed using contrast vectors [Christensen, 2002] that select the regressors of interest for each model, including the temporal and dispersion derivative regressors (Table 3.1). The criteria used to evaluate the models included the F-statistic scores, the number of voxels above the $p < 0.05$ (FWE corrected) and $p < 0.001$ (uncorrected) thresholds, as well as the location of these voxels (inside or outside the 'BOLD activation mask') for each function (Table 3.3).

The transfer functions were compared as follows (a summary of these comparisons can be found in Table 3.1):

i. In order to ascertain whether our main transfer functions showed significant correlations with BOLD, as suggested by the results from the literature on which these functions were based (see Transfer Functions section), we correlated the Total Power, Heuristic and Frequency Response (3 bands) models individually with BOLD, as described above.

ii. Subsequently, we compared the frequency-dependent functions (Heuristic and Frequency Response) with the main frequency-independent function, Total Power. We implemented two pair-wise comparisons (a) Total Power versus Heuristic and (b) Total Power versus Frequency Response (3 bands), which allowed us to probe whether the link between BOLD and neuronal activity is frequency-dependent.

iii. We then implemented a three-way comparison (Total Power, Frequency Response with 3, 5 or 8 bands and Heuristic) to finally determine which transfer function provides a better fit to the BOLD data, when all models are taken into account.

iv. We also performed a similar comparison but we've included the GFP transfer function together with the previous models. This allowed us to assess whether a model based on the amplitude of the EEG signal, rather than its spectral content, was a better predictor of BOLD.

v. To determine whether the Frequency Response model performs better with fewer frequency bands, in particular with just a single band, we've performed two pairwise comparisons between (a) the Heuristic and the power in the Alpha frequencies (8 to 15 Hz) and between (b) the Heuristic and the power in the high (Beta/Gamma) frequency band (15 to 40 Hz).

vi. Finally, to investigate different properties of the Heuristic model, as described above, we implemented several pair-wise comparisons. These included the Heuristic versus (a) the u-Heuristic, (b) the l-Heuristic, (c) the MF function and (d) the Frequency Response model with 5 and 8 bands constructed using the normalised power

spectrum.

For each of the above comparisons we used a fixed effects group analysis using 3-sessions of data from three subjects (9 sessions in total), giving rise to a total of 900 scans. Subsequent analyses based on 1800 scans (200 scans per session, as mentioned above) produced very similar results. These fixed effects SPMs summarise the results over the three subjects [Penny and Holmes, 2006]. We also computed SPMs for each subject in isolation, in a case study approach (see below).

The total number of regressors for each of the design matrices used is summarised in Table 3.1. For example, for the main three-way comparison (iii.) the design matrix employed 198 regressors (198 = 3 regressors of interest for Total Power, 3 for the Heuristic, 9 for the three-band Frequency Response model, 6 for the movement regressors and 1 for the session mean $\times$ 9 sessions).

The stimulus onset-based regressors were not included in these design matrices, since they do not provide a plausible biological model, or link, between BOLD and underlying neuronal activity. These regressors were only used in a preliminary analysis to define a 'BOLD activation mask'.

3.3 Results

3.3.1 Artefact correction and SVD

To remove scanning artefacts from the EEG, the data were processed as described in Section 3.2.4. Figure 3.1 shows the first 10 sec of an example time-series from corrected EEG data for (a) the mean of electrodes O1 and O2 and (b) the scalp signal, $\tilde{y}$, obtained from the SVD. As can be seen, the data appears uncontaminated by MR-related artefacts and is relatively free from other artefacts, such as eye-blinks. A prominent $\sim$ 10 Hz waveform can also be easily detected in these signals.

The fact that the time courses of these two signals look very similar (Figure 3.1) was expected, since the first principal component of the EEG is primarily driven by activity from posterior regions. This is confirmed by plotting the topography of this component, as shown in Figure 3.2. In addition, the first component explains 67 % of the total variance of the data, which should provide a good representation of EEG activity.

After this step we computed steady state visual evoked responses (SSVERs) to further assess the goodness of the MR-related artefact correction method. These

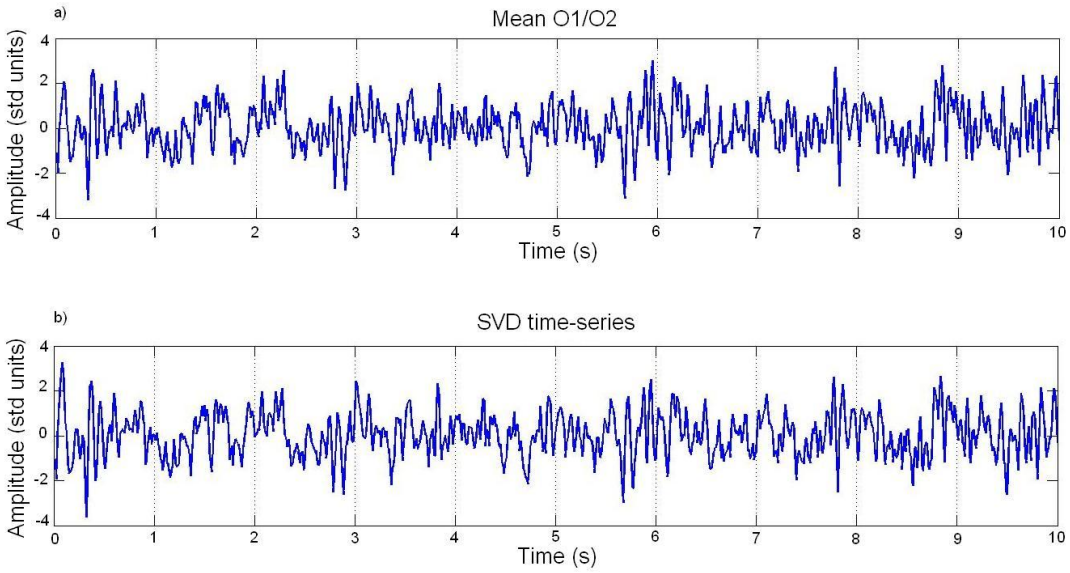

Figure 3.1: Example of artefact corrected EEG time-series for the first 10 seconds of the first visual stimulation period: (a) Mean activity of electrodes O1 and O2. (b) Projection onto first principal component (SVD time-series).

SSVERs were computed by first epoching the artefact-corrected 29-electrode EEG data acquired inside the MRI scanner, for each subject/session, in half-second (500 ms) post-stimulus window and then averaging across trials. Spectral analysis was then performed on the epoched and averaged EEG, using the data from electrode O2 (8 averaged epoch time-series corresponding to the different stimuli used). The time-frequency spectra were constructed using wavelets, as previously described in Section 3.2.4 (Eq. 3.1 and Eq. 3.2). The same procedure was then performed to obtain the SSVERs for the EEG data acquired outside the MRI scanner with the same experimental conditions, including the same paradigm. The responses obtained for both datasets were then compared. Figure 3.3 shows the averaged SSVERs over all sessions of one representative subject for different frequencies of visual flicker.

As can be seen in Figure 3.3 the major component of the spectra is at the second harmonic of the stimulus frequency. This result was expected since for reversing stimuli the SSVERs are usually produced at the phase-reversal or alternation frequency, which is twice the stimulation frequency [Burkitt et al., 2000]. This fact also explains why almost no response is seen for the 30 Hz stimulus, for the range of frequencies here analysed (1 to 40 Hz).

However, for the purpose of this section we were only interested in the similarity

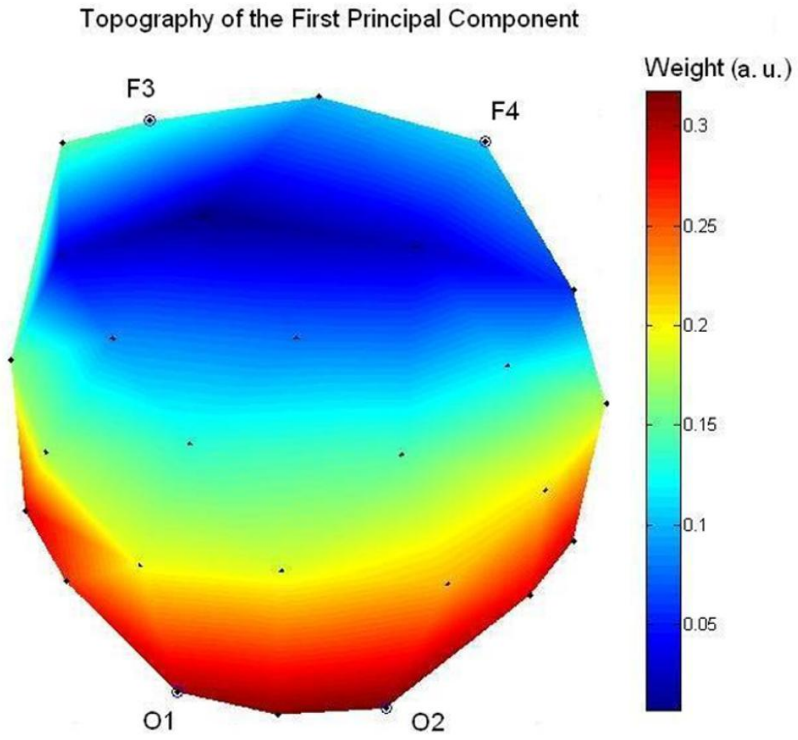

Figure 3.2: Topography (2D) of the EEG first principal component for a representative subject. The locations of the occipital and frontal electrodes are indicated by their respective names.

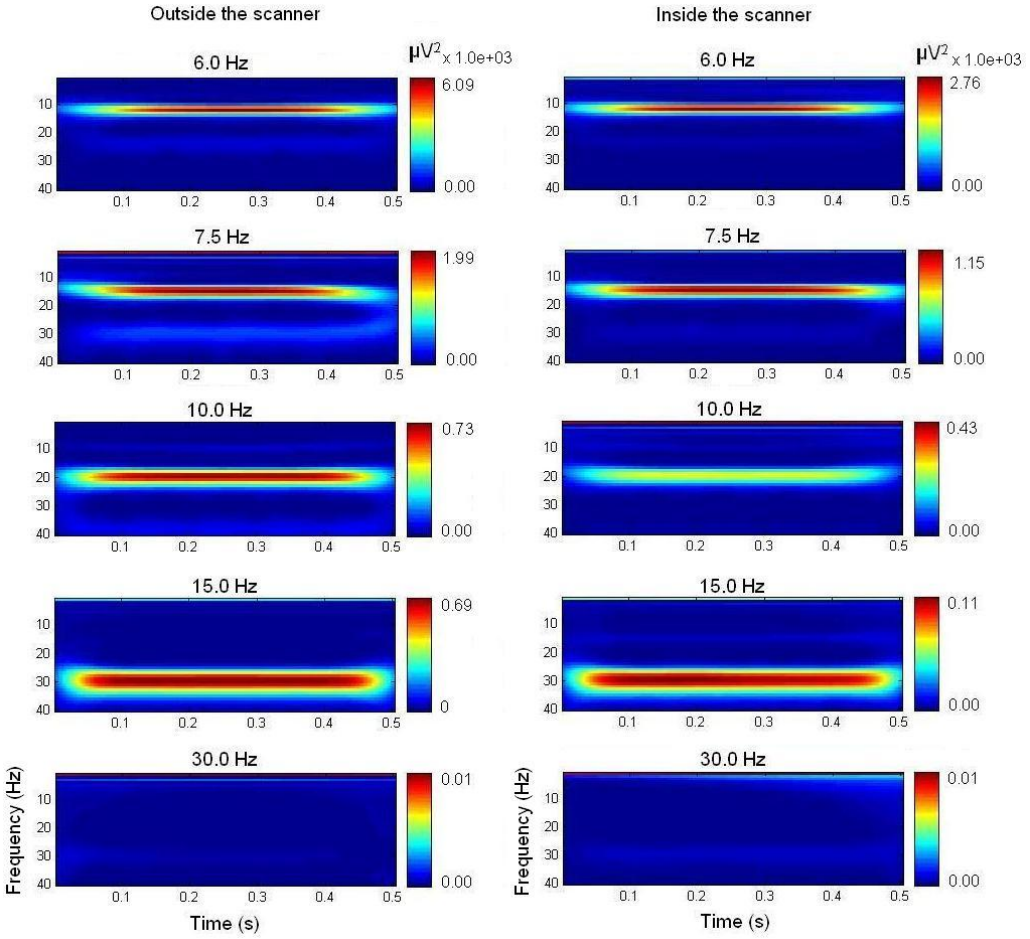

Figure 3.3: Power spectra of the SSVERs for EEG acquired outside (left) and inside the scanner (right) averaged over the three sessions of one representative subject. The frequencies on top of each plot correspond to the stimulation frequencies of the visual flicker stimuli.

between the responses obtained inside and outside the scanner, and as can be seen in Figure 3.3 the close correspondence indicates that the MRI artefacts can be removed without filtering out the signal of interest.

The SSVERs were not used in the subsequent regression analysis. To compare the different transfer functions we used the EEG artefact corrected, un-averaged projected data onto its first principal component.

3.3.2 Effect of the experimental task

We then looked at the effects of visual flicker on both the EEG and fMRI data. For the EEG data, the SSVER spectra shown in Figure 3.3 provide evidence that visual cortical neurons synchronised their firing to the stimuli, leading to strong EEG responses at the second harmonic of the stimulus frequency.

Table 3.2: Anatomical location in Talairach space of the sites with significant results from the three-way model comparison (fixed effects SPM analysis, without SVC).

Regressors	[x, y, z] (mm)	Location	Inference
Stimuli	[12, -101, 18] [-9, -101, 15] [3, -92, 3]	Right Cuneus Left Superior Occipital Gyrus Right Calcarine Gyrus	$p < 0.05$ (FWE)
Heuristic	[-6, -77, 15] [3, -92, 10] [-54, -17, 9] [60, -11, 15] [-12, -62, -12]	Left Calcarine Gyrus Right Calcarine Gyrus Left Superior Temporal Gyrus Right Rolandic Operculum Left Cerebellum	$p < 0.05$ (FWE)
Total Power	[-48, -74, 12]	Left Middle Temporal Gyrus	$p < 0.001$ (unc.)
Freq. Response	[-48, -74, 12] [-42, -74, -15]	Left Middle Temporal Gyrus Left Cerebellum	$p < 0.001$ (unc.)

For the fMRI data, both single subject and fixed effects group analyses showed significant bilateral activation ($p < 0.05$ (FWE)) in visual areas of the occipital cortex

(Figure 3.4). These areas were identified with the help of the 'Anatomy Toolbox' for SPM software [Eickhoff et al., 2005]. Talairach coordinates of cluster maxima [x,y,z] mm: Right Cuneus [12, -101, 18], Left Superior Occipital Gyrus [-9, -101, 15], and Right Calcarine Gyrus [3, -92, 3] (Table 3.2). The fMRI images from the group analysis in Figure 3.4 were used to create the BOLD activation mask, so that subsequent analyses could be restricted to BOLD activated regions.

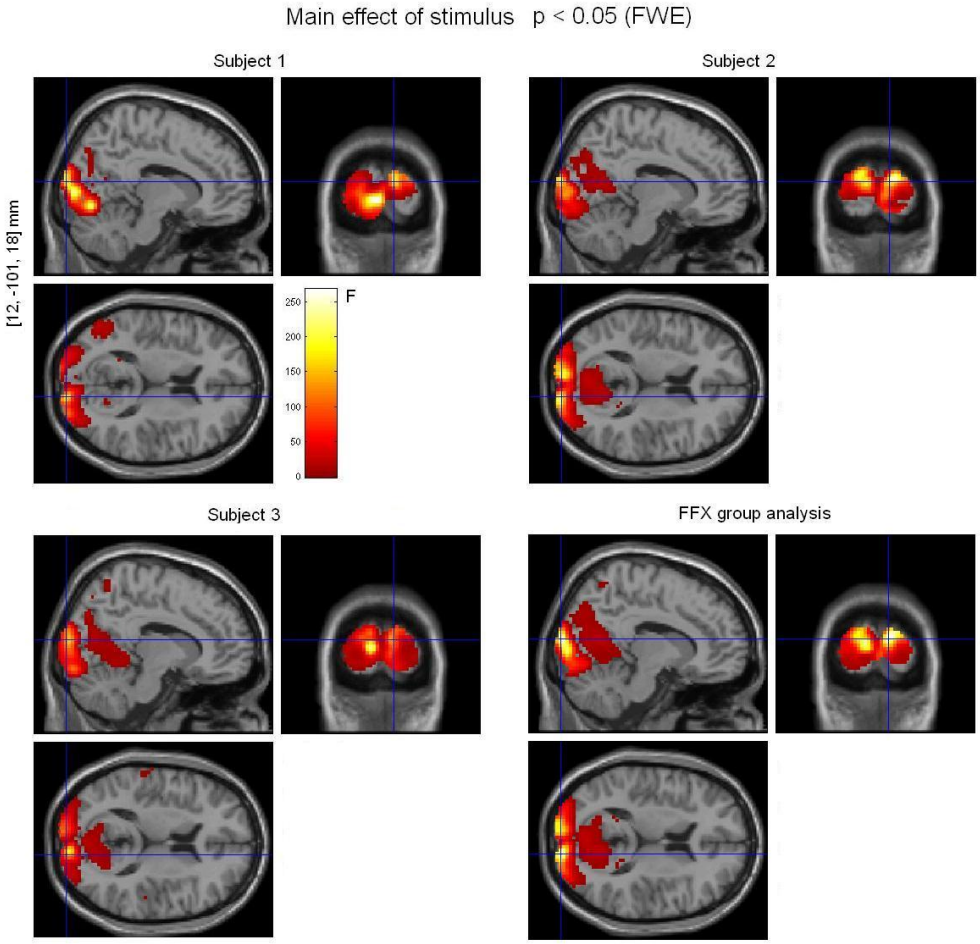

Figure 3.4: Effect of visual flicker stimulation on fMRI data. Single-subject analyses (3 sessions per subject) and Fixed effects group analysis (9 sessions in total), $p < 0.05$ (FWE). The voxel locations written on the left of each figure correspond to the most significant cluster maximum for the group analysis (Talairach space).

In a separate analysis (not shown) which controlled for variation in luminance levels using an additional regressor of no interest, BOLD activity was shown to have an inverted U-shaped response to flicker frequencies below 8Hz (equivalent to 16

Hz for reversing frequencies, see Chapter 5). The peak response was for a flicker frequency of 4 Hz (8 Hz reversing frequency), agreeing closely with previous studies (Singh et al. [2003], Parkes et al. [2004] and Wan et al. [2006]). The results of this analysis are presented in Chapter 5. This result also explains why the amplitude of the SSVERs plotted in Figure 3.3 decreases with increasing stimulus frequency, for both the responses obtained inside and outside the scanner (see scaling in Figure 3.3).

3.3.3 Relationship between neuronal activity and BOLD

Figure 3.5 plots example regressors for the Total Power, Heuristic and Frequency Response (3 bands) models derived from Equations 3.3, 3.4 and 3.5, convolved with the Hemodynamic Response Function and downsampled to the fMRI frequency (1/3.06 sec). Figure 3.5d plots an example BOLD time-series for the same time interval and subject, at the most significant cluster maximum from Figure 3.4 (fixed effects group analysis), in relation to the paradigm. As can be seen there is an increase in BOLD during the 'Task' blocks which is better reflected in the Heuristic than in the other models. The highest frequency band of the Frequency Response model (Figure 3.5c, black) also seems to follow BOLD more closely than the time-series from the other bands.

The SPM analyses with the separate design matrices (one for each model) showed significant ($p < 0.05$ (FWE)) correlations between each model and the observed BOLD signal, as can be seen in Figure 3.6. The locations of maximal correlation for each model were not far apart and were included in the voxels activated by the experimental task shown in Figure 3.4. Although all functions correlated with BOLD, the Heuristic produced higher maximal F scores and more voxels above the chosen threshold ($p < 0.05$ (FWE)) than the other two models (Figure 3.6).

The contrast estimates for the most significant voxel for each model showed that the Heuristic correlates positively with the amplitude of the BOLD response, while Total Power and the first frequency band of the 3-band Frequency Response model correlated negatively with this response (Figure 3.7). Other sites showed significant correlation between BOLD and the other two frequency bands (not shown).

We then performed two pair-wise comparisons (a) between Total Power and Heuristic and (b) between Total Power and the Frequency Response model (Figure 3.8). We included the regressors for the two functions we were interested in comparing in the same design matrix. The results clearly revealed that the Heuristic

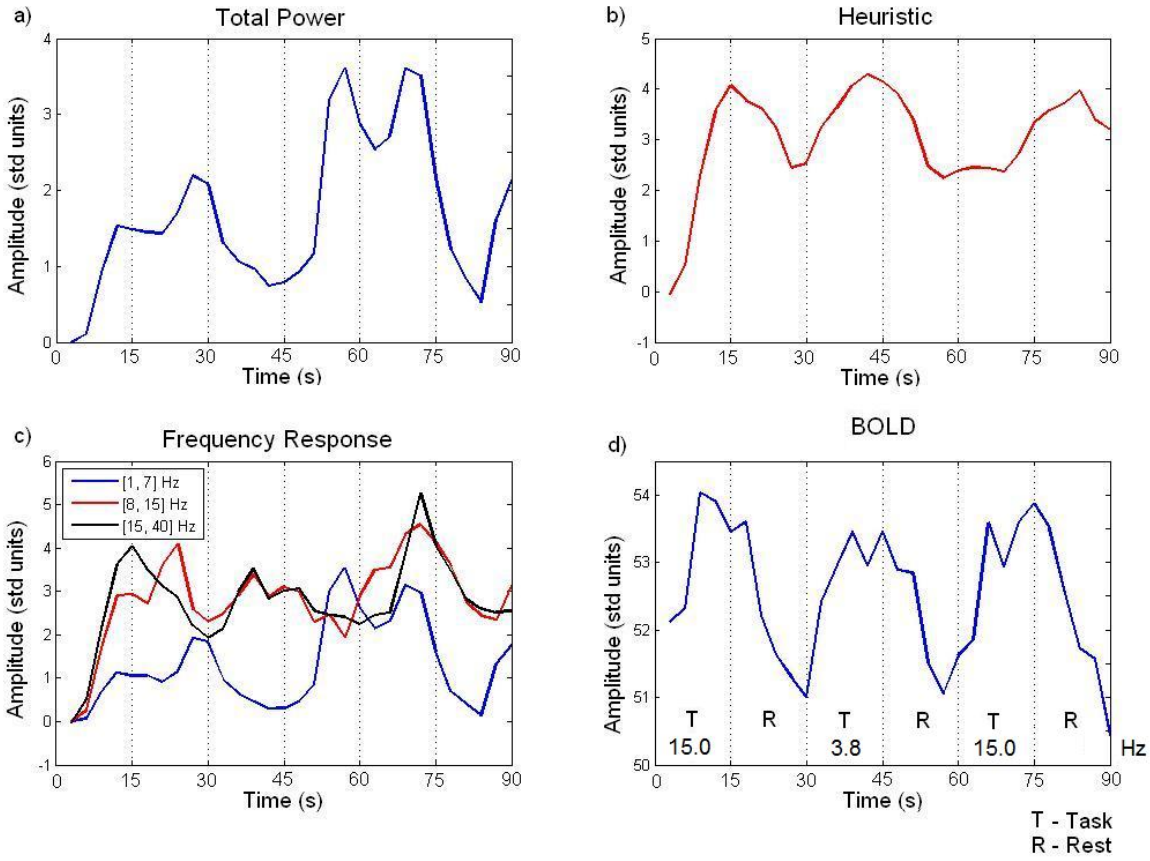

Figure 3.5: Example regressors for (a) Total Power, (b) Heuristic and (c) Frequency Response (3 bands) models after convolution with the HRF (subject 2). (d) Example BOLD time-series for the same period of time and subject, at the most significant cluster maximum ([12, -101, 18] mm, Talairach space) from the fixed effects group analysis of the main effects of visual stimulation (Figure 3.4). The numbers at the bottom of (d) denote the stimulus frequencies.

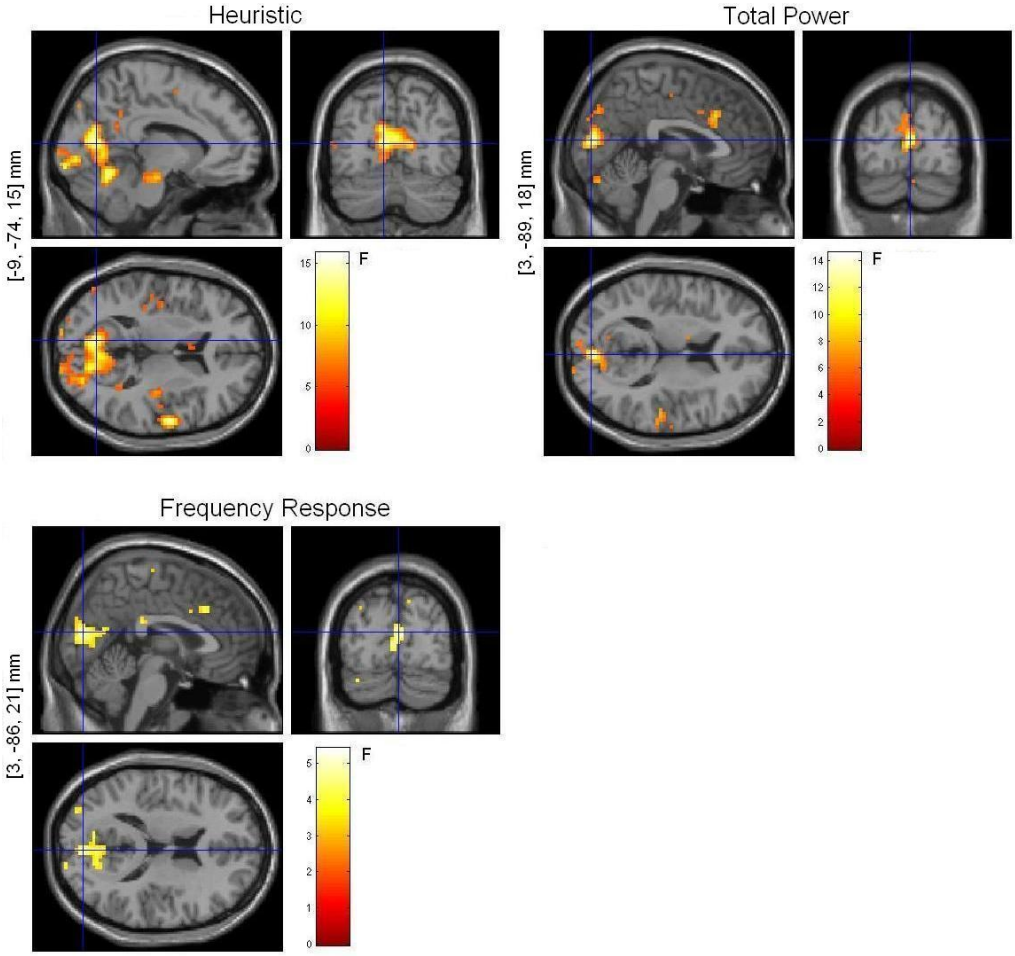

Figure 3.6: Fixed effects SPM analyses ($p < 0.001$ (unc.)) for the Heuristic, Total Power and Frequency Response (3 bands) models analysed in separate design matrices. The voxel locations written on the left of each figure correspond to the most significant cluster maximum after small volume correction with the BOLD activation mask (Talairach space).

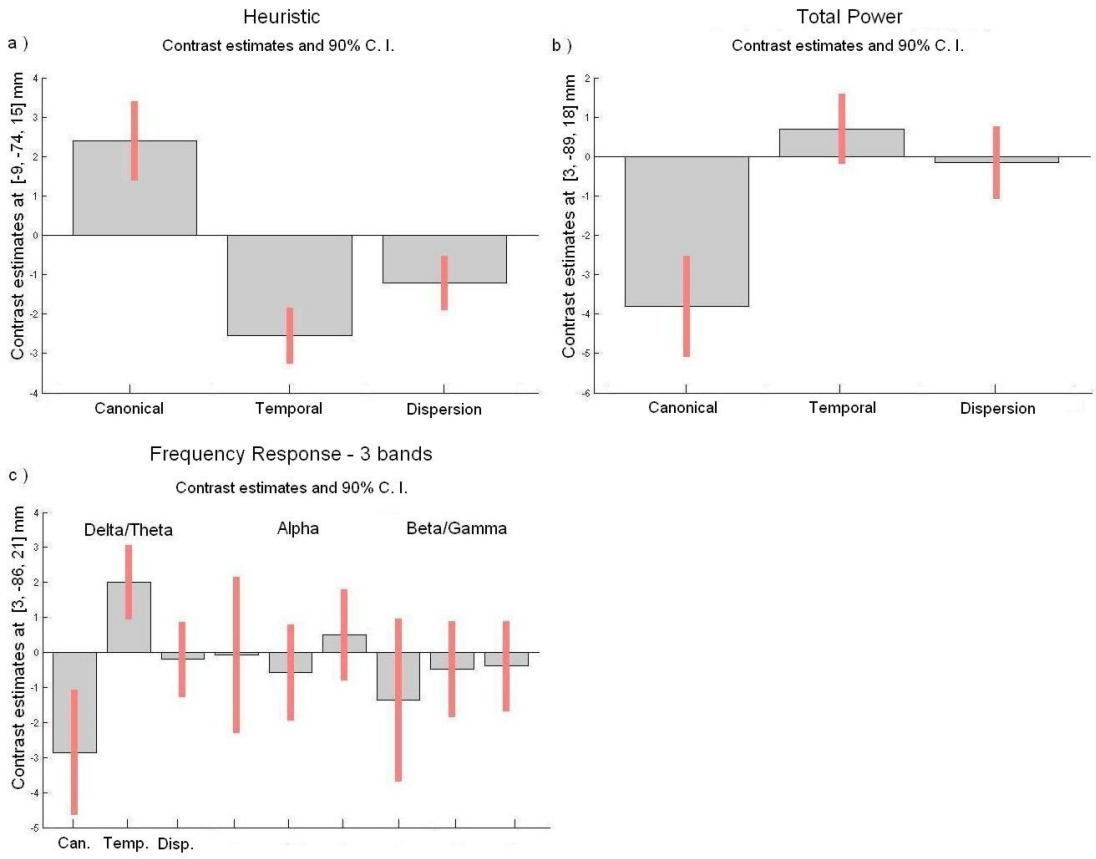

Figure 3.7: Contrast estimates and 90% C.I. for (a) Heuristic, (b) Total Power, and (c) Frequency Response with 3 bands (analysed individually). The estimates include the canonical HRF, as well as its temporal and dispersion derivatives.

provides a much better fit to the data than Total Power. For the second comparison it was difficult to see the effects of each model, since the regressors for the Total Power and particularly the first band from the Frequency Response function (3 bands) were highly correlated.

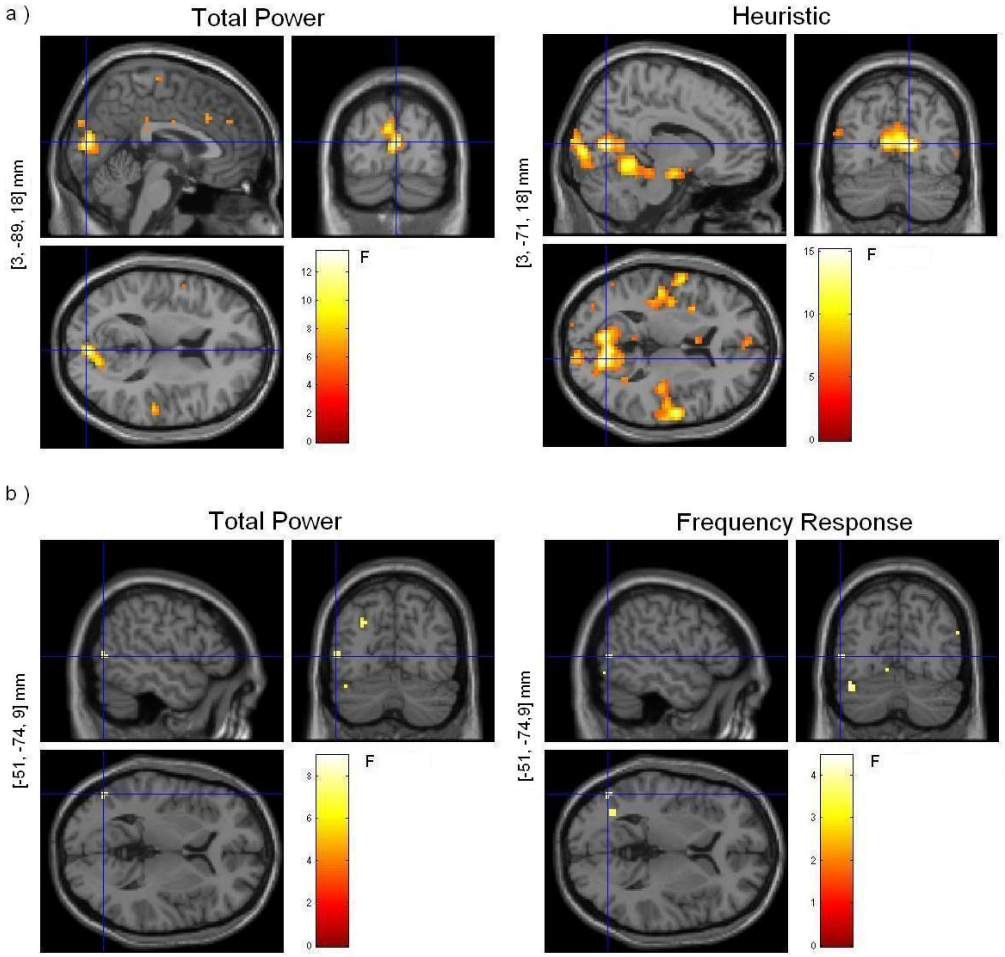

Figure 3.8: Two-way model comparison between (a) Total Power versus Heuristic and (b) Total Power versus Frequency Response (fixed effects SPM analyses ($p < 0.001$ (unc.)). The voxel locations written on the left of each figure correspond to the most significant cluster maximum after small volume correction with the BOLD activation mask (Talairach space). These F-maps show the correlations between EEG and BOLD that are uniquely attributable to each model within a pair.

The three-way comparison, using regressors from all models in the same design matrix, showed a much more widespread and stronger relationship between the Heuristic regressors and the BOLD signal, compared to the Total Power or the Fre-

quency Response functions, $p < 0.001$ (unc.) (Figure 3.9). Furthermore, only the Heuristic showed significant correlations when we corrected for multiple comparisons, ($p < 0.05$ (FWE), using a small volume correction (SVC) (see Chapter 1) over the BOLD activation mask), and the clusters that remained after SVC were located in the Right and Left Calcarine Gyrus (Talairach coordinates [x,y,z] mm: [3, -92, 10] and [-6, -77, 15], respectively) and in the Left Cerebellum (Talairach coordinates [x,y,z]: [-12, -62, -12]) (Table 3.2).

These results are summarised in Table 3.3, where the number of voxels and the highest F-scores obtained for each model, within and outside the activation mask, for different thresholds can be found. The number of voxels, as well as the F-statistics, in both locations and thresholds were significantly higher for the Heuristic than for the other models (Table 3.3).

Table 3.3: Summary of results for the three-way comparison between Total Power (TP), the Heuristic and the 3-band Frequency Response (FR) models from the fixed effects group analysis (Figure 3.9). 'BAM' is the Brain Activation Mask obtained from the main effects of stimulation (Figure 3.4); n_{vox} is the total number of voxels within a region and F_{max} the maximum F-statistic within that region.

		n_{vox} | F_{max}		
Location	Threshold	Heuristic	TP	FR_3
Within BAM				
	$p < 0.05$ (FWE)	17 | 13.3	0 | -	0 | -
	$p < 0.001$ (unc.)	620 | 13.3	5 | 8.2	18 | 4.9
Outside BAM				
	$p < 0.05$ (FWE)	7 | 13.3	0 | -	0 | -
	$p < 0.001$ (unc.)	801 | 13.3	46 | 9.6	95 | 4.8

This three-way comparison is the main result of this work and it was replicated in a case study analysis [Penny and Holmes, 2006] in which data from the different subjects was analysed separately. The individual results were very consistent across subjects: the Heuristic model was markedly superior for all three subjects (individual SPMs not shown), by producing higher F-scores than the rest of the models and more activated voxels inside and outside the Brain Activation Mask. These results are summarised in Table 3.3 (individual tables not shown).

These results were also reproduced when we analysed 1800 scans instead of 900 (see above). Moreover, the inclusion of more data produced even higher statistics

and more significant voxels (in the same brain areas reported) for the Heuristic than for the other models (not shown).

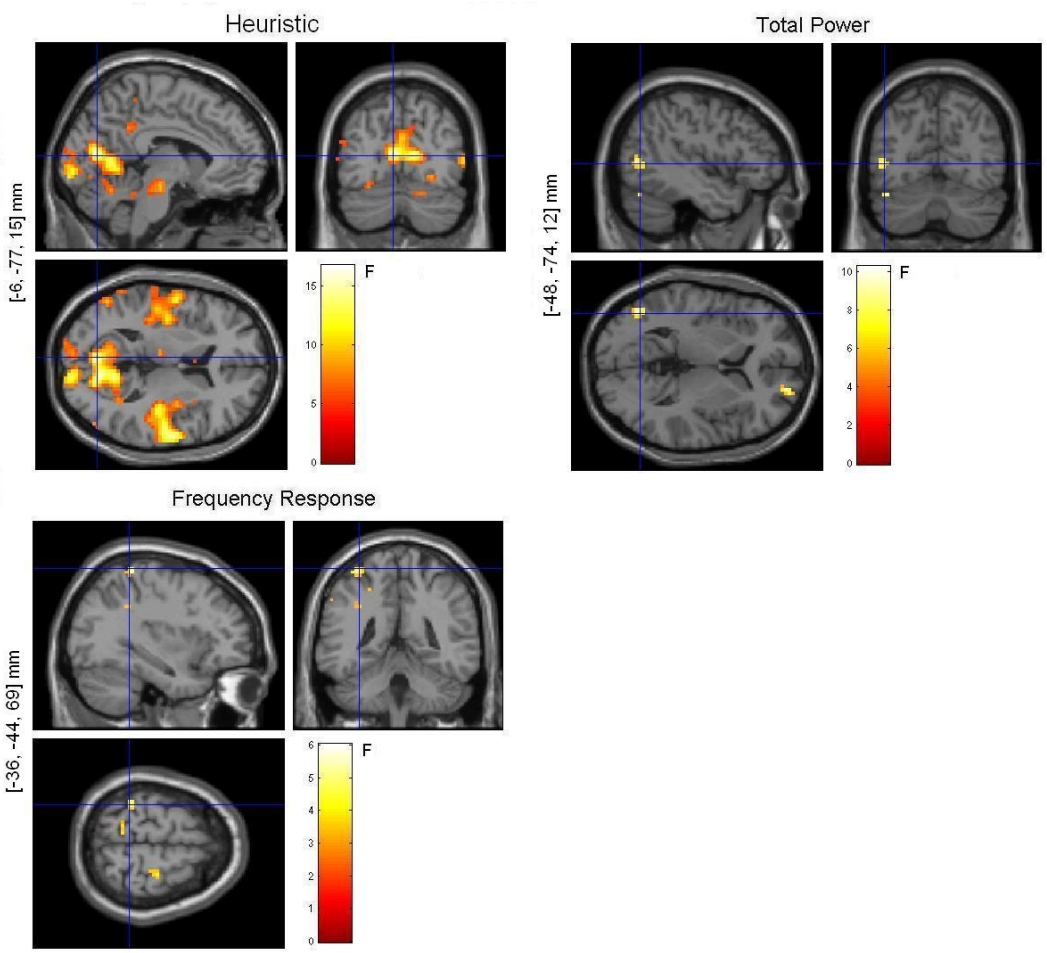

Figure 3.9: Three-way model comparison: fixed effects SPM analyses ($p < 0.001$ (unc.)). Heuristic, Total Power and Frequency Response (3 bands). The voxel locations written on the left of each figure correspond to the most significant cluster maximum after small volume correction with the BOLD activation mask (Talairach space). These F-maps show correlations between EEG and BOLD that are uniquely attributable to each model.

We also compared our three main models (Heuristic, Total Power and Frequency Response (3 bands)) with the Global Field Power of the EEG signal as described in Section 3.2.5. Therefore we added this model to our fixed-effects design matrix. However, the inclusion of this function did not affect the previously obtained results (maps not shown), and the Heuristic again provided a better fit to the data, by pro-

ducing more spatially distributed significant activations ($p < 0.05$, FWE corrected) and higher F scores than the other models, including the GFP. These comparisons allowed us to reject the hypothesis that a model based purely on variations of amplitude across the EEG channels could provide a better fit to the BOLD data.

In addition, when we compared the Heuristic with the single-band Frequency Response models, the Heuristic also revealed more significant voxels and higher statistics than the Alpha and Beta/Gamma power. Moreover, inside the Brain Activation Mask, the number of voxels where the Heuristic provided a better fit (FWE corrected) was 939 (maximum F-statistic: 23.9) when compared with Alpha, and 1480 (max-F: 31.8) when compared with Beta/Gamma. These two models showed only 69 (max-F: 16.8) and 293 (max-F: 20.9) activated voxels in this region, respectively. This result showed that reducing the number of bands in the Frequency Response model didn't improve the performance of this model when compared to the Heuristic (this test was included as it was thought the Frequency Response model might be over-parameterised).

As an aside, we note that although the fMRI data were slice time corrected, significant variability was explained by the temporal derivative regressors (SPMs not shown), and therefore their inclusion in data analyses such as these is recommended (see for example Figure 3.7a). See also recent work by Sladky et al. [2011].

Comparing the Heuristic model and its un-normalised version, the u-Heuristic, revealed that the Heuristic significantly correlated ($p < 0.05$ (FWE)) with the observed BOLD data in most of the brain areas revealed when this function was compared to the Total Power and the Frequency Response models (Figure 3.10). Applying the BOLD activation mask showed that the site with the most significant result was located again in the Right Calcarine Gyrus (Talairach coordinates [x,y,z] mm: [15, -80, 15], p = 1.71e-09 (FWE), SVC) (Table 3.2). In this area BOLD correlated positively with the Heuristic, and negatively with u-Heuristic.

Finally we looked at the importance of the non-linearity present in the RMSF function for the Heuristic model, introduced by the square root operator (the R in RMSF). This was addressed by performing the following two-way model comparisons: between (i) the Heuristic and its linear version, the l-Heuristic (Eq. 3.7), (ii) the Heuristic and the Frequency Response model but using normalised power (eight bands of 5 Hz each) and (iii) the Heuristic and the Mean Frequency function (Eq. 3.8). The rationale behind the second comparison is that the Frequency Response model based on normalised rather than un-normalised power should be able to implement the

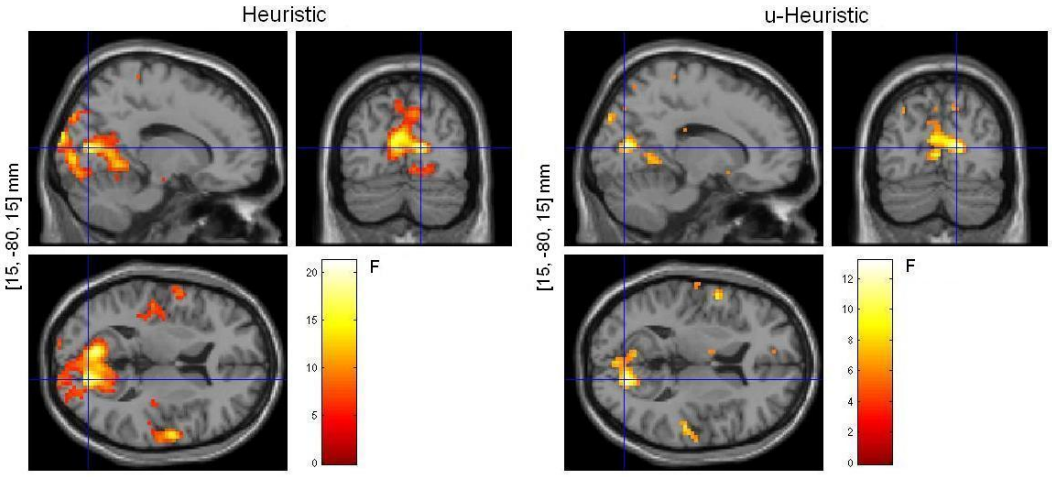

Figure 3.10: Comparison between Heuristic and its un-normalised version, the u-Heuristic: fixed effects SPM analysis ($p < 0.001$ (unc.)). The voxel locations written on the left of each figure correspond to the most significant cluster maximum after small volume correction with the BOLD activation mask (Talairach space). These F-maps show correlations between EEG and BOLD that are uniquely attributable to each model.

transfer function by assigning regression coefficients, $\beta_f = f^2$. The results from these comparisons (SPMs not shown) were very similar. Although when analysed separately all these functions correlate significantly with the BOLD data at a high statistical threshold ($p < 0.05$ (FWE)), when put in the same design matrix none of the models is able to uniquely explain significant variation in BOLD. These results indicate that the nonlinearity introduced by the square root function is not critical.

3.4 Discussion

In this chapter we have used simultaneously acquired EEG and fMRI data, with a visual flicker stimulation task, to probe the transfer function from neuronal activity to BOLD. We compared three different models, each assuming BOLD is sensitive to a different feature of the EEG. These were (i) the Total Power model (ii) the Frequency Response model and (iii) the Heuristic model. When analysed in separate design matrices all transfer functions correlated with the observed BOLD data, as expected.

For the Frequency Response model all bands showed significant correlations with

the data, in agreement with recent monkey EEG-fMRI results [Goense and Logothetis, 2008].

One initially surprising finding was that, at the location of maximal correlation, Total Power correlated negatively rather than positively with BOLD. However, this can be understood by noting that most of the power in the EEG signal, over rest and stimulus blocks, lies in the lower frequencies of the spectrum. This was confirmed by the negative correlation found in the lowest frequency band of the 3-band Frequency Response model in agreement with Mukamel et al. [2005] and Laufs et al. [2006a]. Work in which positive correlation was observed, for example Wan et al. [2006], focussed rather on event-related power (rather than power in both rest and stimulus blocks). In addition, the fact that we modelled the relation between neuronal activity and BOLD in both stimulus and rest blocks together, implies that the Heuristic is also applicable to spontaneous neural activity.

The results of the two-way model comparison, between Total Power and the Heuristic, showed that the transfer function from neuronal activity to BOLD is frequency dependent. The three-way comparison, was again clearly in favour of the Heuristic which was shown to explain significantly more BOLD activity than the other two models.

Independent of model, the majority of the voxels that were significantly correlated with the regressors were in the occipital cortex (Figure 3.6). This is not surprising as we used flickering visual stimuli. What is perhaps surprising is that other brain areas outside of the occipital cortex (such as the cerebellum and temporal cortex) were also significantly correlated with some of the regressors, most notably for the Heuristic model (Figure 3.9). It should be noted that as the Heuristic is a function of the power spectrum and is not a function of any one particular frequency, it may capture some dynamics that are not a simple entrainment of neural populations at some harmonic of the flicker rate.

One concern we had regarding the two and three-way comparison results was that the Heuristic may be better than the Frequency Response model simply because of the small number (three) of frequency bands used. However, our conclusions remained unchanged for frequency response models with additional numbers of bands (five and eight). Conversely, one might also think that the Frequency Response model could do better with a smaller number of frequency bands. The limiting case of this is a single frequency band. Two-way model comparisons, however, revealed the Heuristic to be better than using either (8-15Hz) alpha or (15-40Hz) high (beta/gamma) power

alone.

Our attention then turned to what it is about the Heuristic that makes it a good model. We first addressed the issue of power normalisation. Comparison with a 'scaled' Heuristic, based on un-normalised rather than normalised spectra, revealed the original Heuristic to be clearly superior. The use of normalised power therefore seems important.

We then addressed the issue of nonlinearity. This derives from the square root operator in Eq. 3.5 (the R in RMS). A direct comparison of the Heuristic with its linear version based on the MSF, as well as the Heuristic and the MF model, showed that when included together in the same design, the predictive power of both functions were reduced by the other. Similarly, model comparison of a normalised Frequency Response model with the Heuristic revealed that neither model showed superior predictive power. These results together indicate that, empirically, the non-linearity introduced by the square root function does not appear to be critical. A caveat however is that this conclusion may only be valid for the range of frequencies generated in this experiment (1 to 40Hz).

In the longer term, we envisage that 'asymmetric' [Kilner et al., 2005] regression approaches will be superceded by 'symmetric' forward models, such as proposed in Sotero and Trujillo-Barreto [2008]. Interestingly, this forward modelling approach based on neural mass models also supports the Heuristic, as exogenous input causes both a BOLD activation and an increase in the mean LFP frequency [Sotero and Trujillo-Barreto, 2008].

Some results in the literature may appear at odds with the Heuristic. For instance the positive correlations with alpha power found in the thalamus by Goldman et al. [2002] and in other regions [Gonçalves et al., 2006]. However, the Heuristic describes a relationship based on relative not absolute power. Therefore if increases in alpha were, for example, accompanied by decreases in lower frequencies (delta/theta), this would be compatible with the Heuristic. Using separately acquired fMRI and source-reconstructed MEG data, Muthukumaraswamy and Singh [2008] showed stimulus-related increases in gamma band activity without corresponding changes in BOLD. However, whilst this result clearly speaks against the gamma-BOLD hypothesis, it does not necessarily speak against the Heuristic. This is again because the Heuristic depends on the relative power of the whole spectrum.

An interesting inference to be drawn from Muthukumaraswamy and Singh [2008] is that gamma-band power may reflect the synchronized activity of *local* neuronal

ensembles. This view fits in with neural network modelling results [Kopell et al., 2000] and power-law analyses of electrocorticogram data [Miller et al., 2007]. Whilst BOLD can be sensitive to changes in the gamma band, as many studies have shown, it is also sensitive to activity in the whole spectral domain, including the more spatially dispersed lower frequencies [Kopell et al., 2000], and processes reflecting large-scale neuromodulatory input [Logothetis, 2008].

The original paper that described the Heuristic model was partly inspired by the results of EEG-fMRI integration in the study of epilepsy. In this field, increased slow wave activity has been shown to be associated with decreased BOLD [Archer et al., 2003] while spike and wave discharges (with high frequency components) have been shown to cause BOLD activations [Hamandi et al., 2004; Krakow et al., 2001]. This would be entirely in agreement with the Heuristic model.

To our knowledge our work reports the first study where the model proposed by Kilner et al. [2005] has been empirically tested using human brain imaging data. It is also the first work in which different putative functions for the relationship between BOLD and spectral characteristics of neuronal activity, as measured with EEG, have been explicitly compared.

To this end we designed a study providing experimental control over the frequency structure of the EEG signal by entraining networks to visual stimulation at different frequencies. Our results suggest that changes in BOLD are indeed associated with changes in the spectral profile of the underlying neuronal activity, and that these changes don't arise from a single spectral band. Instead they result from the dynamics of the various frequency components together, in particular, the relative contribution of high and low frequencies as proposed in Kilner et al. [2005].

Although we entrained networks to visual stimulation we have no reason to anticipate different results if neuronal activity were modulated by different cognitive processes. However, this is an empirical question that should be addressed in future studies. The current chapter provides evidence in favour of the Heuristic model but, of course, as with any scientific experiment does not prove that the underlying theory is true. We expect that as data is gathered from additional experimental paradigms and sensory modalities a balance of evidence will emerge.

Understanding the nature of the link between neuronal activity and BOLD plays a crucial role in improving the interpretability of BOLD imaging, and relating electrical and hemodynamic measures of human brain function. Finding the optimal transfer function should also aid the design of more robust and realistic models for the integra-

tion of EEG and fMRI, leading to estimates of neuronal activity with higher spatial and temporal resolution, than are currently available. In Chapter 5 we revisit the same data used in this chapter with a biophysically realistic model of neurovascular coupling to investigate how BOLD relates to synaptic and spiking activity. Before that we switch topics to consider Bayesian model selection maps for group studies. In future, we plan to apply the approach described in the following chapter to evaluate neurovascular coupling models, such as the functions described here.

Chapter 4

Bayesian model selection maps for group studies

4.1 Introduction

Given a set of candidate hypotheses, or models, scientists can use Bayesian inference to update their beliefs about the respective hypotheses, in light of new experimental data. The most likely hypothesis can then be identified using Bayesian model selection (BMS) (see Chapter 2).

BMS is based on the model evidence, i.e. the probability of obtaining observed data, y, given model m, $p(y|m)$. In a group study, one obtains a separate evidence value for each model and for each subject. Under the assumption that the data is independent from subject to subject, these evidence values can be multiplied together to produce a single evidence value for each model. The ratio of resulting model evidences then forms what is known as the Group Bayes Factor [Stephan and Penny, 2007].

In more recent work, Stephan et al. [2009] have shown that the Group Bayes Factor approach corresponds to what is more generally known as a Fixed Effects analysis [Penny and Holmes, 2006] (Chapter 2). A drawback of the FFX approach is that it does not account for between-subject variability which can make the resulting inferences over-confident. Additionally, it is not robust to the presence of outliers.

Stephan et al. [2009] contrast the FFX approach with a proposed Random Effects (RFX) approach, in which a (potentially different) model is assigned to each member of the group. Stephan et al. [2009] then describe Bayesian estimation procedures

for obtaining the posterior distribution $p(r|Y)$, where Y comprises data from all subjects. Contrary to the FFX approach, this method correctly takes into account the variability between subjects, and is also robust to outliers.

In earlier work, Penny et al. [2007] have developed Bayesian spatio-temporal models for fMRI data, which provide within-subject model evidence maps. Voxel-wise comparison of these maps allows neuroimagers to make inferences about regionally specific effects. These comparisons are analogous to the F-tests used in statistical parametric mapping (SPM), with the advantage that the models to be compared do not need to be nested. Additionally, an arbitrary number of models can be compared together.

The Bayesian approach is useful when there is no natural nesting of hypotheses. A trend in recent neuroimaging research, for example, is to fit computational models to behavioural data, and then to use variables from these data fits as regressors in general linear models of fMRI data [Behrens et al., 2008; Montague et al., 2004]. A natural extension of this approach is to derive different sets of regressors from different computational models, and so allow fMRI to provide evidence in favour of one model or another. An example in the field of behavioural control would be to compare different models of 'value updating', e.g. the Rescorla-Wagner model versus the Temporal Difference model [Montague et al., 2004].

In this chapter we describe the combination of the mapping approach for providing log-evidence maps for each model and subject, with the RFX approach described in Stephan et al. [2009] (see Section 2.3.4.2). This procedure constructs posterior probability maps (PPMs) for BMS inference at the group level. We illustrate the method using fMRI data from a group of subjects performing a cued two-choice reaction time task, and compare it with a FFX analysis of the same data. It is our long term aim to apply the methods described in this chapter to adjudicate among models of neurovascular coupling (see Chapters 3 and 5). What follows in this chapter, however, provides a proof of concept that is based on a study of information theoretic models of brain responses.

These models can be used together with brain imaging data to understand how different information-theoretic quantities are learned and represented in the brain, such as entropy and surprise. This is done by relating changes in these quantities, as determined by a particular model, to changes in the observed brain signals. For example, Bestmann et al. [2008], used a cued visual task, to show that motor system responsiveness, measured with transcranial magnetic stimulation (TMS), varies according

to the trial-by-trial entropy (average uncertainty) and surprise (the stimulus-bound information of a visual cue) conveyed by visual events guiding an action.

This chapter is structured as follows. In the next section (4.2) we describe how BMS maps can be constructed from previously estimated log-evidence maps, as well as the data and models we are interested in comparing. In the Results section (4.3) we apply this method to fMRI group data from a cued target detection task, and in Section 4.4 we discuss these results.

4.2 Materials and methods

4.2.1 Subjects and task

Twelve subjects responded to a right or left sided target ("+ O" or "O +") appearing for 250 ms on a screen by spatially compatible button presses using the right and left index finger, respectively. The target was preceded by a visuospatial cue ("< + <" or "> + >") presented for 250 ms and appearing 1000 ms before the target. Four different event types were presented randomly: validly cued right and left button presses (66 trials each), and invalidly cued right and left button presses (17 trials each). During null events (165 trials), the central fixation cross was maintained with no presentation of cue or target, and no corresponding button press. The inter-trial interval was 2000 ms. This experimental paradigm, which involves simple spatially congruous and incongruous cues, is known as a cued detection or Posner-type task [Posner, 1980]. Responses were recorded by computer using COGENT Cognitive Interface Software (Wellcome Trust Centre for Neuroimaging, London, UK).

4.2.2 fMRI acquisition and analysis

fMRI data were recorded using a Siemens VISION system (Siemens, Erlangen, Germany), operating at 2T. A total of 330 functional volumes (28 slices) were recorded for each subject, using T2*-weighted MRI transverse echo-planar images (EPI) (64×64 matrix, $3 \times 3 \times 5$ mm^3 voxel size, TE = 40ms) with blood oxygenation level dependent (BOLD) contrast. Effective repetition time (TR) per volume was 2.15 seconds.

Data were preprocessed using Statistical Parametric Mapping (SPM5, Wellcome Trust Centre for Neuroimaging, http://www.fil.ion.ucl.ac.uk/spm/) implemented in Matlab 6 (The Mathworks Inc., USA). Functional volumes were realigned and un-

warped [Andersson et al., 2001] and the resulting volumes were normalised to a standard EPI template based on the Montreal Neurological Institute (MNI) reference brain in Talairach space [Talairach and Tournoux, 1988] and resampled to $3 \times 3 \times 3$ mm voxels. The time series in each voxel were high pass filtered at 1/128 Hz to remove low frequency confounds and scaled to a grand mean of 100 over voxels and scans within each session.

4.2.3 Models

This section describes the models that were compared using BMS maps. We started by comparing two simple nested models and followed with a comparison of two non-nested information theoretic models.

4.2.3.1 Nested models

We first specified a 'Validity' model (model1), which only includes a column of 1's for the session mean and additional regressors for validly and invalidly cued trials. The paradigm used here is a two-choice reaction time task, and therefore reaction times vary along the experiment, i.e. reaction times are faster when visual information indicates in advance which action the subject will have to make [Bestmann et al., 2008; Hyman, 1953]. These variability should be accounted for and included in the model. For this reason, the two cued trial regressors were parametrically modulated by reaction times[1].

Second, we specified a 'Null' model (model 2) comprising a single column for the session mean. Comparison of these two models could therefore be implemented using a standard F-test approach with classical SPMs, because model 2 is nested within model 1. More generally, however, the BMS approach does not require the models to be nested (see below).

4.2.3.2 Non-nested models

In Bayesian inference, in principle, there is no upper bound on the number of models that can be compared; however, for the purpose of this work, we focus on two

[1]The stimulus functions used to construct the design matrix in a GLM can be modulated by parametric variates (this can be time or trial-specific variates like reaction time) modelling the interaction between the trial and the variate. The events can be modulated by zero or more parameters (for further information see Buchel et al. [1996]).

alternative non-nested models.

Previous work has shown that the effect of the history of past-events on brain activations can be formalized using information theory, as proposed by Strange et al. [2005] and Harrison et al. [2006], under ideal observer assumptions. One finding of these studies was that activity in a wide-spread fronto-parietal network, including bilateral fusiform, parietal, lateral and medial premotor and inferior frontal regions, as well as in bilateral thalamus relates to the surprise conveyed by a trial event.This activation pattern is similar to the task-related activity shown by our 'Validity' model. The surprise inherent in an event (e.g. an infrequently occurring invalidly cued trial) is based on the probability of that event, given previous trials. Here we consider two different ways of estimating these probabilities: i) an ideal observer (Bayes optimal) model and ii) using a finite length window and weighting of events. Although we assume here that the ideal observer is the optimal model this might not always be the case, and the window model may be preferable in certain situations, e.g. in the case of drifting distributions.

The ideal observer and window model are the two information theoretic models that we compare using BMS maps for group studies. For the first model, we calculated surprise from posterior probabilities updated on a trial-by-trial basis using Bayes rule (see Strange et al. [2005] and Mars et al. [2008] for further details). This was then used to predict neuronal responses measured in our fMRI experiment. More specifically, we modeled the onsets of trials with a stick function that was parametrically modulated by the surprise on a given trial. We refer to this model as the 'Ideal Observer' model, as it is assumed that subjects update their beliefs about events in an ideal, or Bayes optimal, manner.

Alternatively, one can relax the assumption that participants are ideal observers. One could, for example, compare a number of models in which the duration and rate of decay with which past observations (trials) are weighted are differently parameterized. For illustrating the BMS approach, we here focus on one case only, in which only a window of data comprising the four most recent trials was taken into account for computing surprise (see Bestmann et al. [2008] for details). We refer to this model as the 'Window' model. This model is suboptimal from an information theoretic perspective because the observer fails to properly accumulate the evidence available within a block. However, as the brain also has other criteria to optimise (e.g. energy use, speed of response), it could be that imaging data provide evidence for it.

4.2.4 Bayesian model selection maps

Each model was estimated with SPM5, using the first-level Bayesian estimation procedure described in Chapter 2, Section 2.2. This produced a voxel-wise whole brain log-model evidence map for every subject and model estimated (see left panel of Figure 4.1). These maps were then smoothed with an 8 mm half width Gaussian kernel.

This section describes how these log-evidence maps are produced and how BMS can be applied voxel-wise, in order to construct posterior probability maps and exceedance probability maps (EPMs) for Bayesian inference at the group level.

4.2.4.1 Within-subject maps

In earlier work, Penny et al. [2005] developed a Bayesian spatio-temporal model for fMRI data, which allows inferences to be made about regionally specific effects using posterior probability maps (PPMs). Similar approaches have been developed previously by Hartvig and Jensen [2000] and Woolrich et al. [2004b]. PPMs represent images of the probability that a contrast of parameter estimates exceeds some specified threshold and their construction has previously been described in Friston and Penny [2003].

The model developed by Penny et al. [2005] extends previous Bayesian modelling approaches for fMRI [Friston et al., 2002a,b] by, among other things, introducing a spatial prior on the regression coefficients. This prior embodies the knowledge that activations are spatially contiguous, and results in an ability to detect more subtle activations. Although this spatial prior was initially two-dimensional (limited to voxels contained in the same slice) this work has since been extended to three-dimensional priors [Harrison et al., 2008].

In more recent work, Penny et al. [2007] have shown how the model evidence can be used to construct within-subject PPMs for model selection. As compared to model comparison based on F-tests using classical inference, this approach has the advantage of allowing the comparison of non-nested models. Additionally it allows for the simultaneous comparison of an arbitrary number of models. As compared to earlier work [Friston and Penny, 2003] based on PPMs of effect size, the approach is advantageous in not requiring an effect size threshold.

In this chapter we have combined the mapping approach used in Penny et al. [2007] to provide log-evidence maps for each model and subject, with the RFX approach

described in Stephan et al. [2009], in order to produce group maps for model selection.

4.2.4.2 Group maps

Once the log-evidence maps have been estimated for each subject and model, as described above, it is possible to construct between-subject posterior probability maps, that enable inference on model space at the group level. These maps are created by applying the RFX approach, described in Chapter 2, at every voxel, i, of the log-evidence data, which produces a family of posterior distributions, $p(r_{ki}|Y_i)$. We can then construct the PPMs for each model k by plotting the posterior expectation, $< r_{ki}|Y_i >$ for every voxel i (Eq. 2.27) at which the value exceeds a user-specified threshold, γ.

In addition to the group-level PPMs, the RFX approach also allows the construction of EPMs. These constitute an exceedance probability for each voxel i, φ_{ki}, (see Eq. 2.28) and for each model k. Again, these maps are thresholded at a user-specified value γ.

The maps described here can be constructed as whole brain images or images from selected regions of interest. The latter can be created by specifying a mask image, which limits the construction of the maps to voxels contained in the mask. Such masks can be created, for example, using a functional localiser analysis [Friston et al., 2006]. This restricts the analysis to those parts of the brain that are functionally relevant. The overall approach for creating BMS maps for group studies is shown in Figure 4.1.

It is also possible to create group maps using an FFX rather than the above RFX approach. This is implemented simply by summing the log evidence images over subjects for each model (see Eq. 2.23). Posterior model probabilities are then obtained by exponentiating the resulting sums and normalising to unity.

4.3 Results

This section illustrates the application of Bayesian model selection maps to group fMRI data acquired from subjects performing a cued target detection task, as described above (Section 4.2).

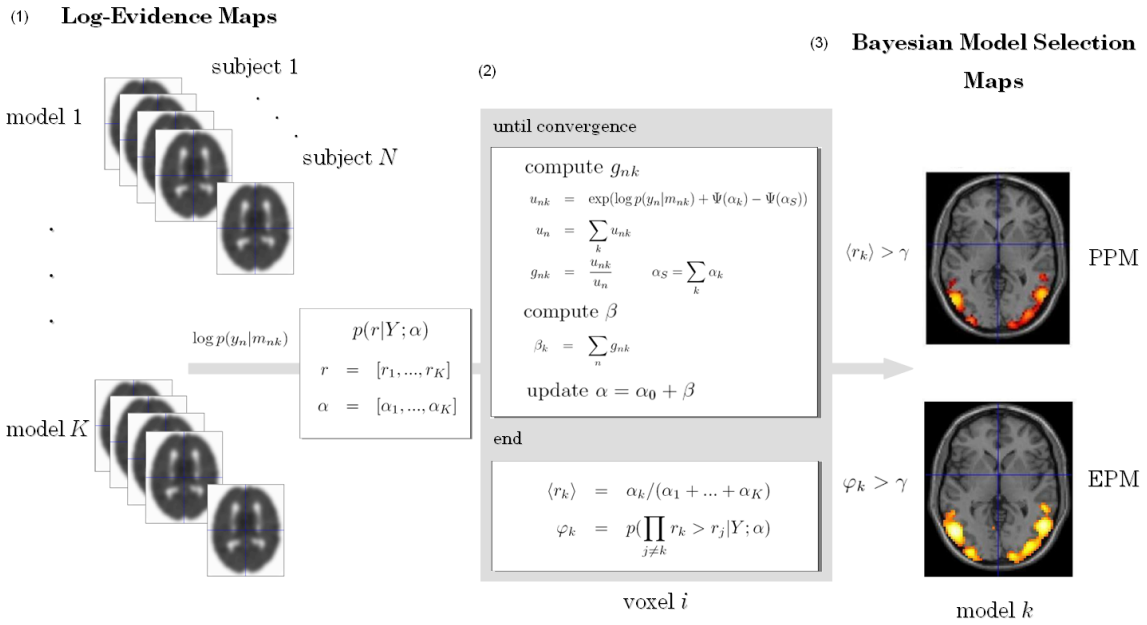

Figure 4.1: Schematic representation of the method for constructing Bayesian Model Selection (BMS) maps for group studies. (1) The first step involves estimating log-evidence maps for each subject and model. (2) The RFX approach for BMS described in the text is then applied in a voxel-wise manner to the log-evidence data. (3) The BMS maps (Posterior Probability Map, PPM; Exceedance Probability Map, EPM) for each model are then constructed by plotting the posterior and exceedance probabilities at each voxel ($< r_{ki} >$ and φ_{ki} respectively), using a threshold, γ, to visualise the resulting image. See the main text for a detailed explanation of the different steps involved in this procedure.

4.3.1 Nested models

To illustrate the construction of BMS maps we started by comparing the two nested models described in the previous section ('Validity' versus 'Null' model).

We applied the RFX approach described in Chapter 2 to the group model evidence data for these two models in a voxel-wise manner. This procedure yielded a PPM and EPM for each model. In addition, we compared these PPMs with those obtained using a FFX analysis.

Figure 4.2 shows the group-level PPMs for the 'Validity' model constructed using the FFX (A) and RFX (B) method, and thresholded in order to show the brain regions where the posterior probability for the 'Validity' model is above $\gamma = 0.75$.

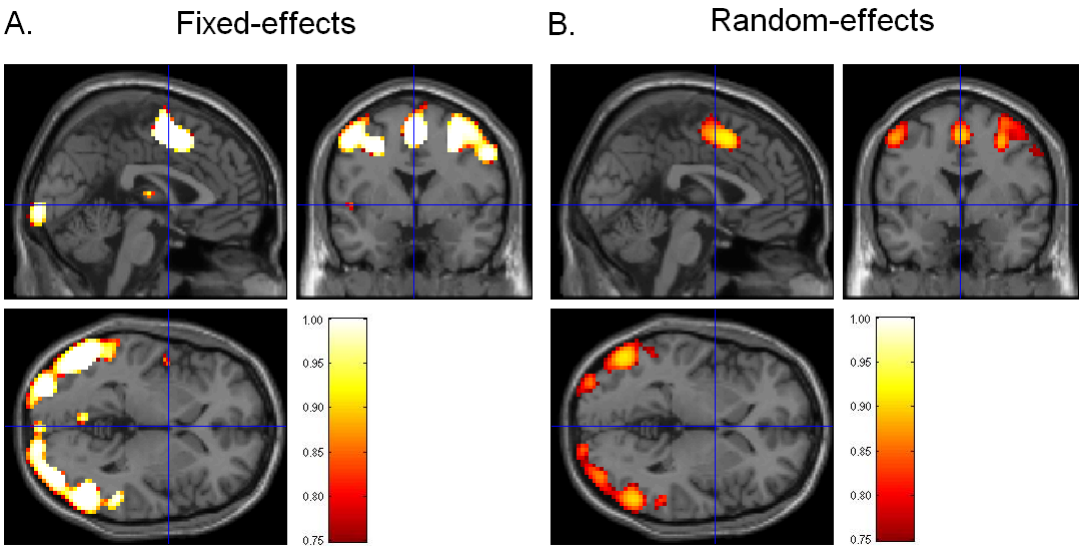

Figure 4.2: Group-level PPMs for the 'Validity' model from (A) Fixed and (B) Random Effects analysis. The maps therefore show brain regions encoding cue validity. These maps were thresholded to show regions where the posterior model probability of the 'Validity' model is greater than $\gamma = 0.75$. The FFX approach does not account for between-subject variability and, consequently, can appear over-confident.

These regions show strong evidence in favour of the 'Validity' model. More specifically, these regions comprise brain areas one would a priori expect to be generally involved in a Posner-type task as used in the example data set presented here [Rounis et al., 2006], including motor areas (peak voxel Talairach coordinates [x, y, z] mm: Left Supplementary Motor Area [0, 5, 56], Right precentral gyrus [33, -4, 53] and Left precentral gyrus [-51, -4, 56]), as well as, visual and attention-related regions

(Talairach coordinates [x, y, z] mm: Right inferior temporal gyrus [57, -67, 2], Left inferior temporal gyrus [-51, -76, 2], Left middle temporal gyrus [-54, -73, 5]). Figure 4.2 shows that the FFX and RFX approaches for inference on model space yielded similar results. However, because the FFX approach does not accommodate between-subject variability the resulting inferences are somewhat over-confident. This is also illustrated in Figure 4.3 where, for example, the position of the crosshairs indicates a cluster that is only visible for the FFX maps.

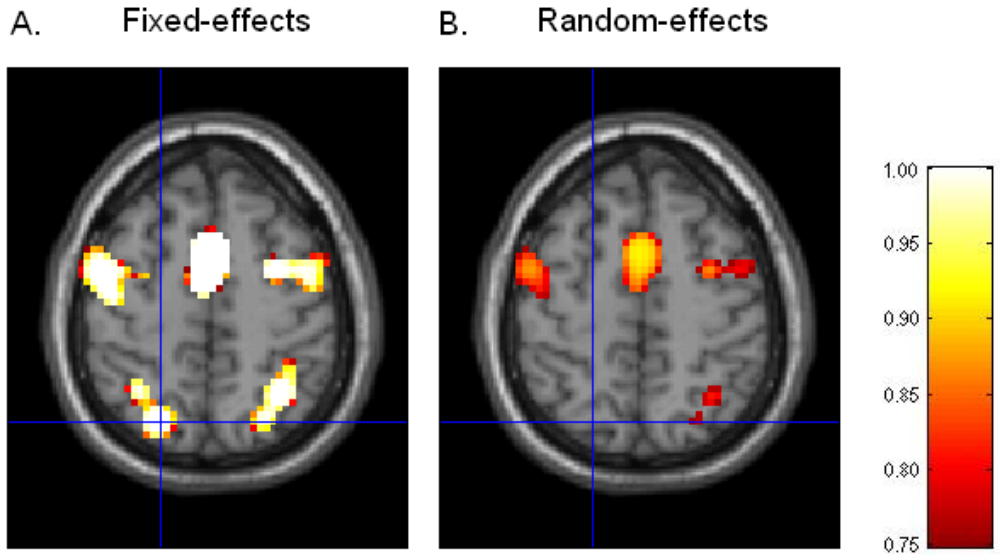

Figure 4.3: Group-level PPMs ($z = 59$mm, Talairach coordinates) for the 'Validity' model from (A) Fixed and (B) Random Effects analysis. The maps were thresholded to show regions where the posterior probability of the 'Validity' model is greater than $\gamma = 0.75$. The position of the crosshairs (Talairach coordinates: [-21, -73, 59] mm) indicates a cluster that is only visible for the FFX maps.

The probabilities obtained for both models at the peak voxel of this cluster are shown in Figure 4.4. As can be seen, the RFX analysis produces lower posterior probabilities for the 'Validity' model than does the FFX approach. Moreover, this probability is approximately 0.7 (Figure 4.4B), which is slightly below the threshold, $\gamma = 0.75$, used for constructing the maps in Figure 4.3. For this reason the corresponding cluster is missing in the RFX map (Figure 4.3B).

Figure 4.5A plots the EPM for the 'Validity' model using a threshold of $\gamma = 0.95$. For this model the exceedance probability is given by $\varphi_{i1} = p(r_{i1} > 0.5)$ and Figure 4.5A plots φ_{i1} only at those voxels for which $\varphi_{i1} > \gamma$. This map is similar to the PPM

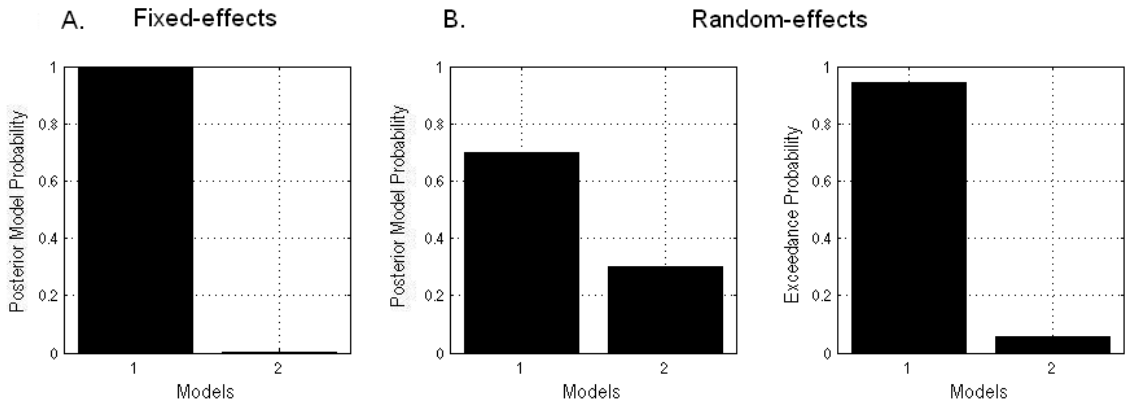

Figure 4.4: Posterior model probabilities obtained by comparing the 'Validity' and 'Null' model (model 1 and 2, respectively) at an example voxel, [-21, -73, 59] mm (Talairach coordinates), using a (A) Fixed and (B) Random Effects analysis. For the RFX analysis we include the exceedance probabilities at the same voxel. As can be seen, the RFX analysis produces lower posterior probabilities for model 1 than does the FFX approach.

shown in Figure 4.2B, which plots $< r_{i1} >$ at those voxels for which $< r_{i1} >> \gamma$.

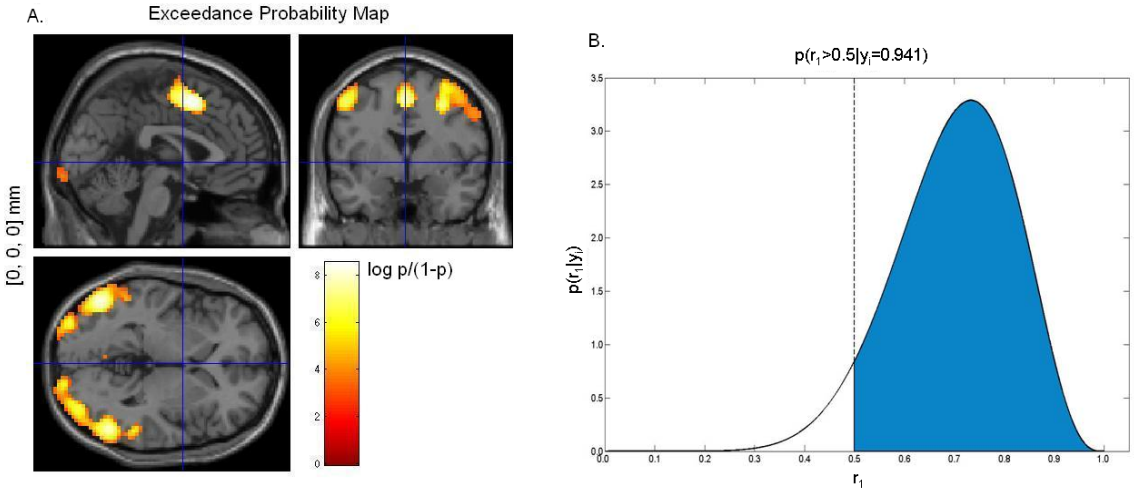

Figure 4.5: (A) Group-level EPM, (log-odds scale) for the 'Validity' model. The map was thresholded to show regions where the exceedance probability for the 'Validity' model is greater than $\gamma = 0.95$. (B) Posterior distribution and exceedance probability for the same model at an example voxel, [-21, -73, 59] mm (Talairach coordinates).

To better illustrate what is being plotted in Figure 4.5A we've plotted the poste-

rior distribution for the same model, $p(r_1|Y)$, obtained at one example voxel (Figure 4.5B). The shaded region corresponds to $r_1 > 0.5$ and for this voxel encompasses 94.1% of the total mass of the posterior distribution. Therefore the exceedance probability value plotted for this voxel is 0.941.

Stephan et al. [2009] have noted that the RFX approach is more robust in the presence of outliers than is the FFX method. We examined this in our data by inspecting regions in the BMS maps showing contradictory results for FFX and RFX. Consequently, we found groups of voxels at which model 1 was clearly the best model for the FFX analysis and model 2 for the RFX. We then looked at the log-model evidence values for all subjects at these voxels and found that the reason for the discrepancy was indeed an outlying subject. Figure 4.6 shows an example of this, where almost all subjects indicate that model 2 is best, except for a single outlying subject with an extreme evidence value favouring model 1.

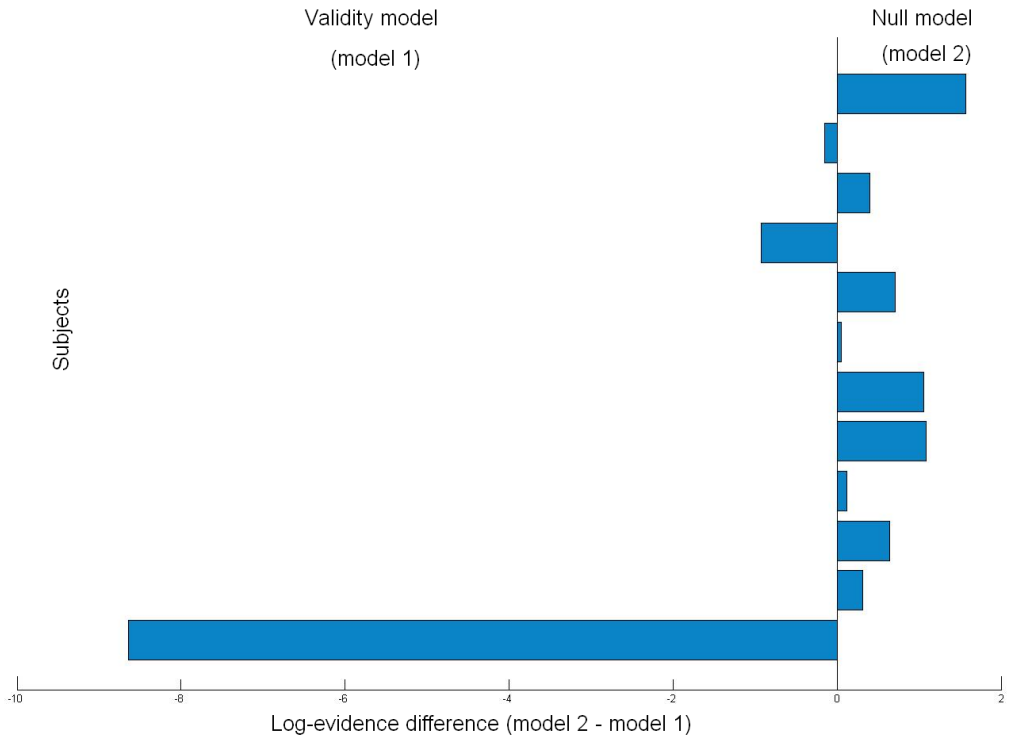

Figure 4.6: Log-model evidence differences between the 'Null' and 'Validity' models (model 2 and model 1, respectively) at voxel [-29, 0, 49] mm (Talairach coordinates), for the 12 subjects analysed. The data clearly show that one subject (bottom row) is an outlier.

The posterior probabilities obtained for this voxel (for which one of the subjects is an outlier) reveal that the FFX results are in favour of the 'Validity' model, while RFX suggests that the 'Null' model is better (Figure 4.7A and B), as can also be seen in the respective PPMs (Figure 4.8). Moreover, the exceedance probability value for the 'Null' model is almost 80%, which indicates strong evidence in favour of model 2 at this voxel.

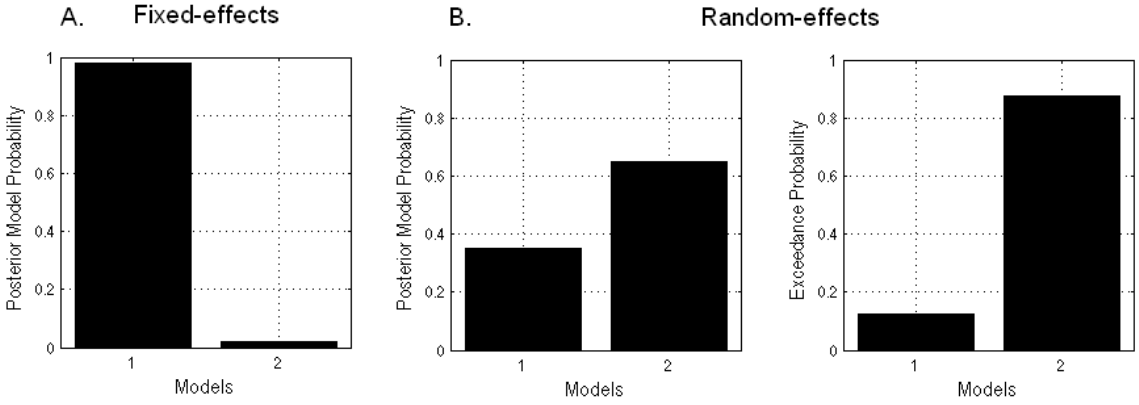

Figure 4.7: Posterior model probabilities obtained by comparing the 'Validity' and 'Null' model (model 1 and 2, respectively) at voxel [-29, 0, 49] mm (Talairach coordinates), using a (A) Fixed and (B) Random Effects analysis. For the RFX analysis we include the exceedance probabilities at the same voxel. The voxel chosen here belongs to a brain region where FFX and RFX analyses yield different results due to the presence of an outlier (see Figure 4.6).

These results corroborate Stephan et al. [2009] who have also shown that the RFX approach is more robust in the presence of outliers.

4.3.2 Non-nested models

The BMS approach presented here is particularly suited for comparing non-nested models. Here, we use the aforementioned example dataset to illustrate how BMS can be applied to compare models for which there is no natural nesting.

These are information theoretic models of brain responses based on probabilistic updates from (i) ideal observer model and (ii) window model.

Figure 4.9a) shows the group-level PPM for the two locations in which the posterior model probability for the 'Ideal Observer' model is greater than $\gamma = 0.6$. We focused explicitly on task-related brain regions, as identified in the group-level PPM

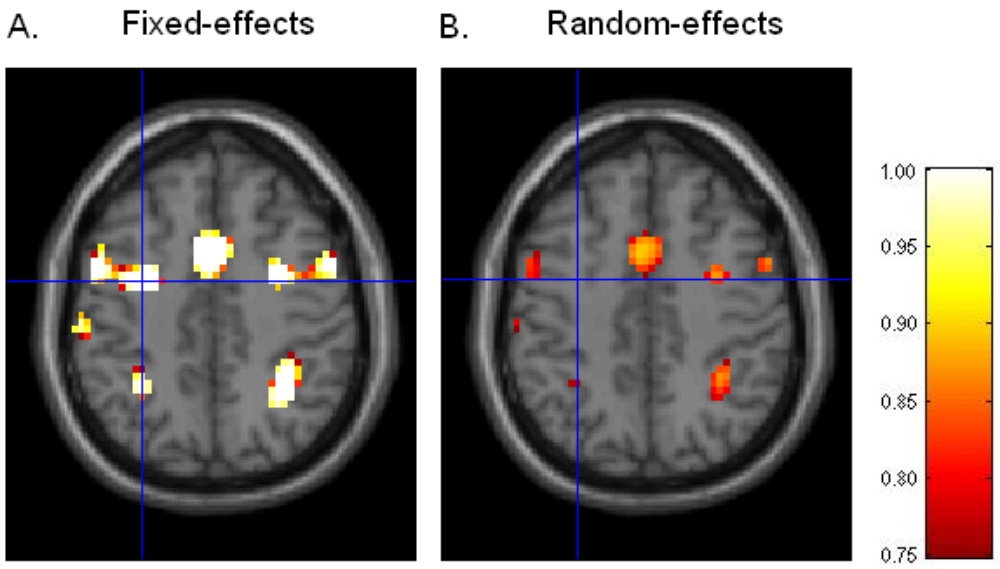

Figure 4.8: Group-level PPMs (slice $z = 49$mm, Talairach coordinates) for the 'Validity' model from (A) Fixed and (B) Random Effects analysis. The maps were thresholded to show regions where the posterior model probability of the 'Validity' model is greater than $\gamma = 0.75$. The crossbars indicate a cluster of voxels where one of the subjects is clearly an outlier (Figure 4.6).

for the 'Validity' model (see Figure 4.2B). This masking approach is an example of a functional localiser (differences in brain activation are only expected for those brain regions engaged by the task). Our BMS suggests that activity in these two regions (Talairach coordinates [x, y, z] mm: Supplementary motor area [6, 5, 56]; Right superior parietal lobule [36, -58, 59]) is best explained by the surprise conveyed by an event, as estimated by an ideal observer.

Figure 4.9b) shows the group-level PPM for the 'Window' model and corresponding task-related regions where the posterior probability for this model is above $\gamma = 0.6$. As can be seen, in other areas, such as the Left Middle Temporal Gyrus, the 'Window' model is a better model than the 'Ideal Observer' to explain the data. Although the main aim of this work was to validate the BMS maps approach, this is an interesting result (Fig. 4.9). This result suggests that different areas of the brain integrate information over different time-scales, with more posterior regions discarding all but the last events in the environment, whilst more anterior regions are concerned with events that occurred over longer times scales. This result has been confirmed in a follow-up study using the BMS maps technique [Harrison et al., 2011].

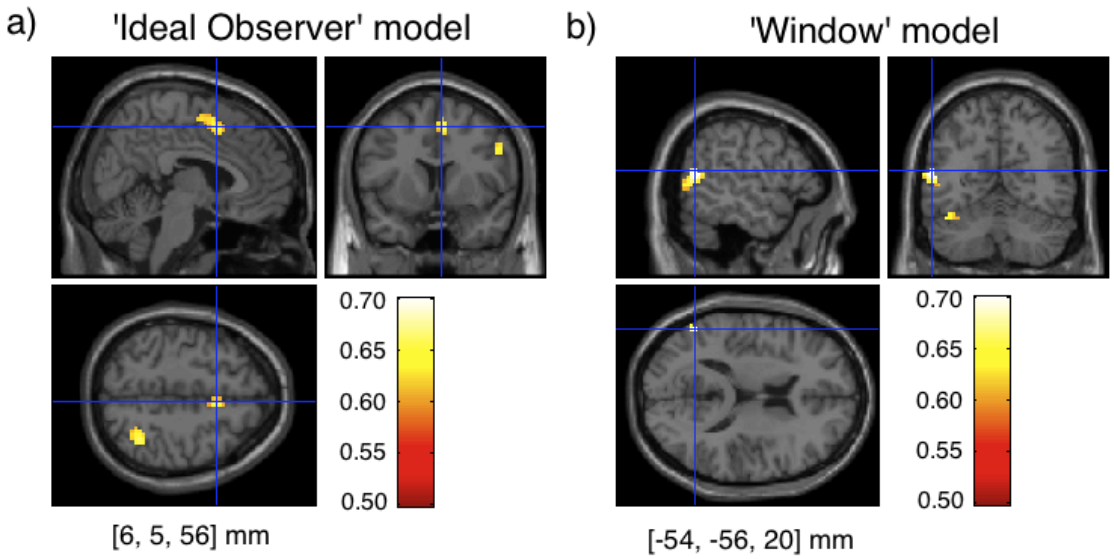

Figure 4.9: Group-level PPMs for: a) the 'Ideal Observer' model and b) 'Window' model from Random Effects analysis. The maps show task-related regions (by masking the results with the 'Validity' model PPM) and are thresholded to show only regions where the posterior model probability for each model is greater than $\gamma = 0.6$. The coordinates written on the bottom of each figure indicate the location of maximum posterior probability.

4.4 Discussion

In this chapter we have presented the construction of posterior probability maps allowing for Bayesian model selection at the group level. These maps are produced by combining a model evidence mapping approach with an RFX approach for model selection.

We have illustrated our method by applying it to fMRI data from a group study, and compared the resulting maps with those obtained using a FFX analysis. As expected, both analyses yielded similar results, but the posterior model probabilities from FFX appeared over-confident. This observation reflects the fact that the RFX inference properly accommodates between-subject variability, whereas FFX does not.

Another important point is the behaviour of the method in the presence of outliers. Since the RFX approach takes into account group heterogeneity, it has proven [Stephan et al., 2009] to be more robust than FFX. In our fMRI analysis we have confirmed this result. Moreover, we have observed that the two analyses yield contradictory results for brain regions where one of the subjects provides strong evidence in favour of one particular model, contrary to the rest of the subjects. The results from FFX are adversely influenced by this single subject, whereas the RFX inference was not.

A minor disadvantage of our new approach is that it relies on the prior computation of log-evidence maps for each subject and model. These computations are more time consuming than the standard SPM approach by a factor of five to ten. However, these individual subject maps need only be computed once for all subsequent group BMS analyses. The method proposed here for constructing BMS maps is not so computationally demanding and takes on average less than half an hour to create whole-brain PPMs for the comparison between two models using the log-evidence images from 12 subjects on a standard PC. Moreover, we envisage that our new approach may be most usefully applied to regions or networks of regions previously identified using functional localiser methods. The use of these localisers has the advantage of speeding up the computation and reducing its time to approximately less than a minute for a region with a few thousand voxels. We also envisage that, in future, subject specific log-evidence maps could be efficiently computed using post-hoc model selection or Savage-Dickey ratios (see Chapter 6).

In this work, log-evidence maps were smoothed by a user specified Gaussian kernel. This will be finessed in future work to include a spatial model over r and its smooth-

ness estimated using a Bayesian framework. This procedure mirrors corresponding developments in the analysis of group data from M/EEG source reconstructions [Litvak and Friston, 2008].

The product of the analysis procedures described in this chapter are posterior probability maps. These show voxels where the posterior probability over model frequency exceeds some user-specified value. In previous work [Friston and Penny, 2003] we have derived PPMs over effect size. We note that, as is common-place in Bayesian inference, these posterior inferences could be augmented with the use of decision theory. This requires the costs of false negative and false positive decisions to be specified. One can then use decision theory to make decisions which minimise, for example, the posterior expected loss [Gelman et al., 1995]. In addition, we note a connection between posterior probabilities and false discovery rate, in which if above threshold values are declared as activations, a posterior probability of greater than 95% implies a rate of false discoveries less than 5% [Friston and Penny, 2003]. It is also possible to relate posterior probabilities to the realised false discovery rate (FDR) (rather than an upper bound or the expected FDR) [Muller et al., 2007]. Finally, we note that a comprehensive Bayesian thresholding approach has been implemented by Woolrich et al. [2005]. This work uses explicit models of the null and alternative hypotheses based on Gaussian and Gamma variates. This requires a further computationally expensive stage of model-fitting, based on spatially regularised discrete Markov Random Fields, but has the benefit that false-positive and true-positive rates can be controlled explicitly.

Unlike classical inference using F-tests, our framework allows for comparison of non-nested models, which we hypothesize will be useful in a number of experimental domains. One such domain is model-based fMRI [O'Doherty et al., 2007] in which computational models are first fitted to behavioural data, and sets of regressors derived to be used as predictors of brain imaging data. A typical example is the study of behavioural control using computational models and fMRI [Montague et al., 2004]. The use of model comparison maps in addition to model-based fMRI would allow brain imaging data to directly adjudicate, for example, between different computation models of value updating [Montague et al., 2004].

In this chapter we focused on information theoretic models of cued detection. Harrison et al. [2011] have recently built on our work and used the framework developed here to adjudicate between windowed updates of event probabilities. The authors compare models with short versus long time windows and find that anterior

brain regions integrate events over larger time windows than posterior brain regions, which is in line with what we observed in this paper. This provides evidence in favour of the hypothesis that the specific time scale which engages a cortical area can be inferred by its location from primary sensory to high level areas [Kiebel et al., 2008].

Chapter 5

Bayesian comparison of neurovascular coupling models using EEG-fMRI

5.1 Introduction

Functional magnetic resonance imaging (fMRI) is an extensively employed neuroimaging technique that allows non-invasive recordings of neuronal activity from the human brain with relatively high spatial resolution. However, the blood oxygenation level-dependent (BOLD) contrast on which fMRI is based is only an indirect measure of this activity. The processes that link underlying neuronal activity to BOLD signals are still a topic of much debate. In particular, there is no consensus on the relative roles of synaptic and spiking activity in the generation of BOLD signals. In order to relate findings from fMRI research to other measures of neuronal activity it is important to understand the underlying neurovascular coupling mechanism [Anonymous, 2009].

Most of our present knowledge about neurovascular coupling comes from animal experiments, combining hemodynamic (e.g. CBF) with electrical measurements (e.g. LFP, MUA), as described in Chapter 1. This body of evidence (Table 1.1) supports the hypothesis that BOLD signals are more closely coupled to synaptic input and processing activity than to the output spikes of a population of neurons [Logothetis, 2008; Thomsen et al., 2004; Viswanathan and Freeman, 2007]. In other words, the BOLD signal is thought to result from pre-synaptic activity and the release of neu-

rotransmitters, in particular glutamate [Bonvento et al., 2002], which then triggers the release of vasodilatory substances, such as nitric oxide, by surrounding glial cells [Aubert and Costalat, 2005; Carmignoto and Gomez-Gonzalo, 2010; Estrada and DeFelipe, 1998; Friston, 2008; Wang et al., 2005].

As discussed in Chapter 1, despite the consensus that is emerging from animal data, the above scenario has been more difficult to assess with human data. One of the reasons is the fact that the participation of healthy subjects prohibits the use of invasive electrophysiological measures, which can only be obtained under very specific circumstances, such as the case of pre-surgical neurophysiological patients (see Chapter 1 for references).

Here we address this issue by providing a modelling framework that can be used to explicitly investigate competing biological mechanisms for the relationship between neuronal activity and the BOLD response in the healthy human brain. Our aim is to explore the relative contribution of synaptic and spiking activity to the generation of fMRI signals in visual cortex.

The framework used here is based on a mathematical model that allows us to non-invasively infer the degree of local synaptic and spiking activity using EEG-fMRI data, in which subjects were exposed to a reversing checkerboard of varying frequencies. This is similar in spirit to the use of 'virtual electrodes' in EEG analysis [Baillet et al., 2001a], but provides more specific biophysical information. This framework consists of a biophysically informed forward model from neuronal activity to the observed EEG and fMRI signals.

Models linking neuronal activity to EEG/MEG signals have been proposed by Jansen and Rit [1995], David and Friston [2003] and Sotero et al. [2007], to mention a few. These models usually use one or two state variables to represent the mean electrical activity of neuronal populations at the macro-column level, and are referred to as neural mass models [Marreiros et al., 2008a].

Models linking 'neuronal activity' to BOLD signals include the metabolic models proposed by Shulman et al. [2001] and Aubert and Costalat [2005] and the Balloon model, proposed by Buxton et al. [1998]. The Balloon model describes how evoked changes in blood flow are transformed into the BOLD response and has been extended by Friston et al. [2000], who introduced a blood flow-inducing signal relating 'neuronal activity' and CBF, and by Sotero and Trujillo-Barreto [2007], where different metabolic pathways have been proposed for supporting excitatory and inhibitory synaptic activity. In the above metabolic and hemodynamic models, stimulus input

functions are used as surrogates for 'neuronal activity'.

Models linking a common underlying neuronal substratum to both EEG and fMRI signals have also been developed [Valdes-Sosa et al., 2009]. Some models are phenomenologically motivated, such as the 'Heuristic' proposed by Kilner et al. [2005], which was investigated in Chapter 3. This model aims to explain empirical results which relate frequency-specific power changes in EEG with fMRI signals and predicts that increases in the BOLD contrast reflect increases in the root mean squared (RMS) frequency of EEG. We have validated these predictions in previous work [Rosa et al., 2010b] using simultaneous EEG-fMRI data in humans with a visual flicker stimulation task (see Chapter 3). As predicted by Kilner et al. [2005], the RMS frequency significantly explained more BOLD activity than the total time-varying spectral power or any linear combination of frequency-band amplitude modulations (e.g. alpha or gamma power). This work is described in more detail in Chapter 3.

Biophysically motivated models include Babajani and Soltanian-Zadeh [2006], Riera et al. [2006], and Sotero and Trujillo-Barreto [2008]. Most of these theoretical frameworks combine the neural mass model approach for EEG with the Balloon model for fMRI, but the coupling between neuronal activity and blood flow differs from model to model. For instance, Babajani and Soltanian-Zadeh [2006] propose that the squared post-synaptic membrane potentials from both excitatory and inhibitory cells from a cortical area drive increases in cerebral blood flow, whilst Sotero and Trujillo-Barreto [2008] consider all the incoming action potentials from populations within and outside the voxel to be the input to the BOLD response. In Riera et al. [2006] this input is proportional to the total concentration of nitric oxide (NO) synthesised by neurons in the cortical unit. The parameters of this NO model have been estimated using EEG-fMRI data from the visual cortex of one subject exposed to a reversing checkerboard with varying frequency [Riera et al., 2007].

Despite these theoretical efforts, the existing modelling frameworks have not yet been used in conjunction with real electrophysiological and hemodynamic data to compare different neurovascular coupling mechanisms, although important steps in this direction have been taken by Riera et al. [2006] and Sotero et al. [2009]. In Sotero et al. [2009], the authors have compared different models to investigate the role of excitatory and inhibitory activity in the generation of BOLD signals, using fMRI data from one subject. They found BOLD signals to be best explained by excitatory activity alone.

In this chapter, we use the model proposed by Riera et al. [2006] and embed

it within a Bayesian framework. Using EEG and fMRI data in combination with Bayesian inference allows us to estimate the underlying synaptic and spiking activity, along with other biophysical model parameters. These quantities are computed using the variational Laplace method described in Friston et al. [2007a]. This optimisation scheme has been successfully applied to other input-state-output systems, such as Friston et al. [2005a] and Moran et al. [2008].

However, inverting generative models using multimodal datasets, can be a technically demanding task, if the temporal characteristics of the datasets are very different, which is the case for EEG-fMRI data. Here we develop a computationally efficient scheme for model inversion. Instead of inverting the model in a single (computationally demanding) step we adopt a 'multi-step inversion' approach. This approach is based on partitioning model inversion into multiple, independent and computationally efficient steps that are motivated by the time-scales of data involved. This is a general procedure that can be used with other datasets and in other multimodal studies, such as with MEG-fMRI or LFP-fMRI data.

Finally, once equipped with this mathematical and computational framework we posit models embodying different hypotheses about neurovascular coupling and adjudicate between them using Bayesian model evidence [Penny et al., 2004]. We compare three models. The first assumes that blood flow depends on the amount of vasodilatory substances (e.g. nitric oxide) released as a result of synaptic activity (*synaptic input model*), as proposed by Riera et al. [2006]. The second assumes blood flow is driven by the firing rate of pyramidal cells from the same unit (*spiking output model*). These hypotheses are then compared against a third model where both these quantities contribute to the BOLD response (*mixture model*). We note that this has nothing to do with Gaussian mixture models from statistics and machine learning.

This chapter is structured as follows. Section 5.2 presents the biophysical model and different neurovascular coupling hypotheses. In Results (Section 5.3), we compare these hypotheses using EEG-fMRI data from visual cortex, whilst in Discussion (Section 5.4) these comparisons are discussed in light of previous findings in the literature. We use the same EEG-fMRI data as was presented in Chapter 3.

5.2 Materials and Methods

5.2.1 Local electro-vascular (LEV) model

We use a realistic biophysical model, proposed by Riera et al. [2006], of how electrical and vascular dynamics are generated within a cortical unit. The unit comprises three subpopulations of cells: two layer IV GABAergic interneuron populations (the transmission and feedback interneurons (INs)) and a layer V pyramidal cell (PC) population (Figure 5.1(a)). Interneurons are modelled as single compartment neurons, whilst each pyramidal cell has three compartments (soma, basal and apical tuft dendrites). With these neural mass formulations, model variables represent the mean of each quantity in each subpopulation, e.g. the mean firing rate of pyramidal cells (see below). Here we briefly describe the forward model. A summary of all the equations and parameters of the model can be found in Appendix A. For a more detailed description please consult the original work [Riera et al., 2006].

5.2.1.1 Neural mass model

A neural mass model (NMM) characterises the population dynamics of electrical states such as the membrane potentials in the somas of the neurons and electric currents flowing in the neuropil. This modelling framework has been introduced in Chapter 1 and is appropriate for data that reflect the behaviour of neuronal populations, such as EEG and fMRI data.

The time variations of membrane potential in the individual compartments of the pyramidal cell and single compartment interneurons, $V(t)$, are determined by the differential equation for a simple voltage source circuit:

$$\tau_m \frac{dV(t)}{dt} + V(t) \quad = \quad I(t)R_m, \tag{5.1}$$

where R_m is the effective membrane resistance of the compartment, and is cell-type and compartment specific. τ_m is the membrane time constant (same for all cells and compartments). The current, $I(t)$, that flows through the membrane of the cell depends on the connections between different elements of the cortical unit and its external inputs (Figure 5.1(a)). The cortical unit receives external excitatory input in different subpopulations, whilst its sole output is the firing rate of the pyramidal

cells, which is a function of I^+ (see below), the current flowing out of the cell. The excitatory inputs to the transmission interneuron, I_3^+, and basal dendrites of the pyramidal cell, I_1^+, correspond to thalamo-cortical afferent projections. The input to the apical tuft dendrites, I_2^+, mediates cortico-cortical interactions. These currents can be found in Figure 5.1(a).

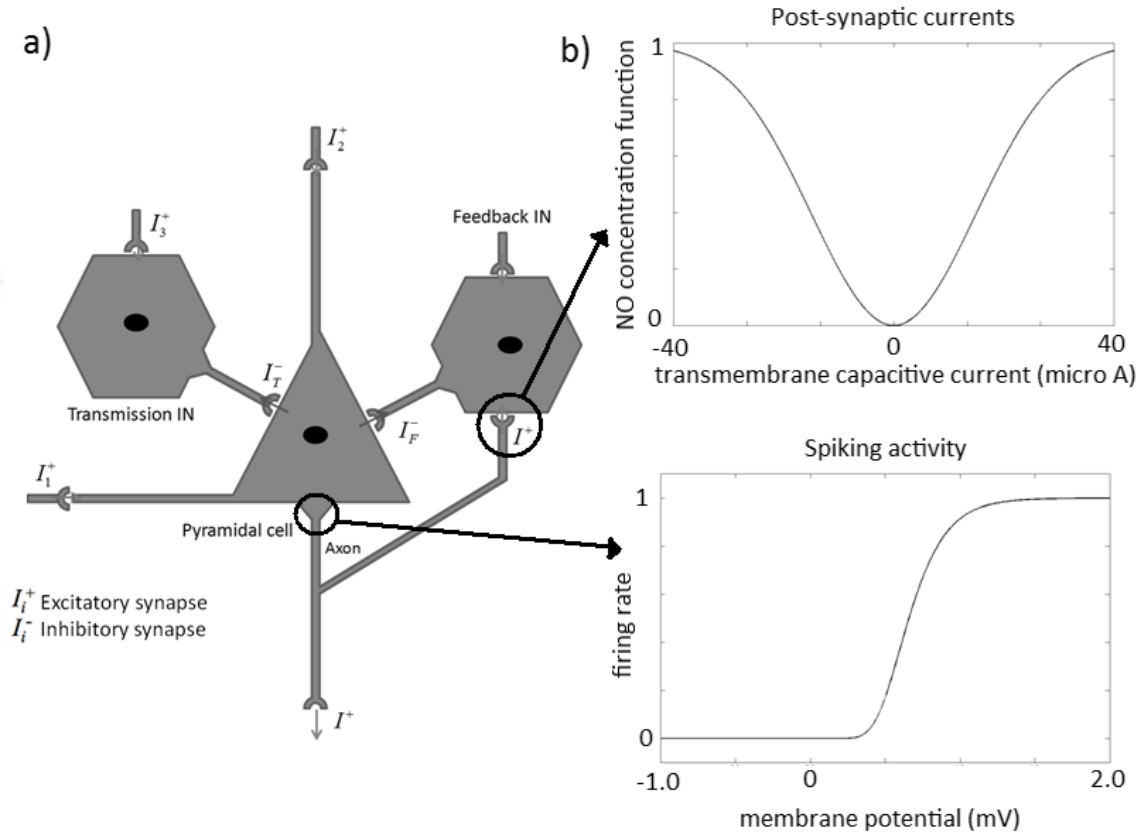

Figure 5.1: Local electro-vascular model: cortical unit. a) The unit comprises three subpopulations of cells, two layer IV GABAergic interneurons and a layer V pyramidal cell. The unit receives input from cortical (I_2^+) or thalamic connections (I_1^+, and I_3^+), whilst its output is the spiking rate of layer V pyramidal cells, I^+; b) Non-linear function of the transmembrane capacitive currents used to calculate the NO concentration. This function is symmetric because both positive and negative currents increase the amount of NO released. This function is used in the synaptic input coupling model. c) Sigmoid function from membrane potential to firing rate. This function is used as the input to the vascular equations in the spiking output model.

In terms of synaptic connections within the cortical unit, the total inhibitory synaptic effect on the pyramidal cell is given by: $I^- = I_T^- + I_F^-$, where I_T^- is the trans-

mission inhibitory current and I_F^- the feedback inhibitory current. The inhibitory synaptic currents depend nonlinearly on the membrane potential of the GABAergic cells through a threshold function: $I^- \propto f(V^{IN})$. The excitatory synaptic current generated by the pyramidal cell has the same form: $I^+ \propto f(V^{PC})$:

$$f(V) = A_l + \frac{A_u}{(1 + Te^{-\gamma(V(t)-V_0)})^{1/T}}. \tag{5.2}$$

The parameters are set to $A_l = 0$ and $A_u = 1$ to ensure that the output stays between 0 and 1. The V_0 and γ parameters determine the voltage sensitivity by setting the membrane potential maximum growth and growth rate, respectively. These parameters are estimated from the data. $T = 0.03$ determines the membrane potential near the asymptote where maximum growth occurs. The threshold function, $f(V)$, is also used to construct the firing rate coupling model (see below).

The equations for the membrane potential at the soma of the three-compartment pyramidal cell, as well as the extracellular potential along its apical dendrites can be determined from the potentials and currents at the individual compartments (given by Eq. 5.1). These equations can be found in Appendix A. The apical dendrites of the layer V pyramidal cells are arranged in parallel to each other and perpendicularly oriented to the surface of the cortex. This geometry facilitates the summation of electric currents in the neuropil. The mesoscopic effect resulting from the spatial average of these extracellular currents corresponds to the electrical signal measured with EEG.

The state variables, $\mathbf{x_N}$, and parameters, θ_N, of the neural mass model described above are summarised in Table 5.1 of the main text and Tables 1 and 2 in Appendix A.

5.2.1.2 Extended Balloon model

The coupling between local neuronal activity, described by the neural mass model, and subsequent changes in vascular dynamics is our question of interest. These changes are expressed in the BOLD signal and have previously been modelled in an extended Balloon approach [Friston et al., 2000], in which a set of four ordinary differential equations comprise the hemodynamic forward model from 'neuronal activity' to hemodynamic responses (see Chapter 1). The full derivation of these equations can be found in Buxton et al. [1998] and Friston et al. [2000]. In brief, for a particular region, neuronal activity, z, causes an increase in a vasodilatory signal, s, that is

subject to auto-regulatory feedback. Inflow, f responds in proportion to this signal with concomitant changes in blood volume ν and deoxyhemoglobin content q (Figure 1.6):

$$\begin{aligned}
\frac{ds(t)}{dt} &= z(t) - \frac{s(t)}{\tau_s} - \frac{f(t)-1}{\tau_f} \\
\frac{df(t)}{dt} &= s(t) \\
\tau_0 \frac{d\nu(t)}{dt} &= f(t) - \nu(t)^{1/\alpha} \\
\tau_0 \frac{dq(t)}{dt} &= \frac{f(t)}{E_0}[1-(1-E_0)^{1/f(t)}] - q(t)\nu(t)^{(1-\alpha)/\alpha}.
\end{aligned} \tag{5.3}$$

The hemodynamic parameters, $\theta_{_H} = \{\tau_s, \tau_f, \tau_0, \alpha, E_0\}$, comprise the rate constant of the vasodilatory signal decay, the rate constant for autoregulatory feedback by blood flow, transit time, Grubb's vessel stiffness exponent, and the resting oxygen extraction fraction, respectively.

The whole dynamic system is driven by the input $z(t)$ (Figure 1.7). Different inputs, z, correspond to different aspects of neuronal activity and consequently different coupling hypotheses between neuronal activity and the BOLD response. A summary of the hemodynamic model's state variables, $\mathbf{x_H} = \{s, f, \nu, q\}$, and parameters, $\theta_{_H}$, can be found in Table 5.1 of the main text and Tables 1 and 2 in Appendix A.

In the next section we specify the neurovascular coupling mechanisms we are interested in comparing.

5.2.1.3 Observation equations

The original electro-vascular model proposed by Riera et al. [2006] is represented by a set of stochastic differential equations describing the dynamics of the neuronal and vascular states, $\mathbf{x}(t)$. In Riera et al. [2006] the stochastic aspect of the model is instantiated by incorporating an additive multidimensional Wiener process to model physiological noise. In this chapter, however, we use a deterministic version of the model. This means that the dynamics are completely determined by the state of the system and stochastic effects enter only at the observation level (Eq. 5.4). This deterministic approach resulted in very similar frequency-response curves to those in Riera et al. [2006] (see Results: synthetic data) and allows us to use standard

Bayesian estimation routines, widely used with deterministic forward models for EEG (e.g. Moran et al. [2008]) and fMRI (e.g. Friston [2002b]).

The observation equations for EEG, $\mathbf{y}_N$, and fMRI, $\mathbf{y}_H$, data are then given by:

$$\begin{aligned} \mathbf{y}_N(t) &= \mathbf{h}_N(\mathbf{x_N}(t)) + \epsilon_t^N \\ \mathbf{y}_H(t) &= \mathbf{h}_H(\mathbf{x_H}(t)) + \epsilon_t^H, \end{aligned} \tag{5.4}$$

where the errors are assumed to be i.i.d., $\epsilon^{N,H} \sim N(0, \sigma_{N,H}\mathbf{I})$.

The temporal variations of the EEG signal are well approximated by the extracellular electric current in the neuropil, $\rho(t)$, obtained from the NMM multiplied by the lead field matrix, $\mathbf{L}$. This matrix contains information about the geometry and conductivity of the head, and is therefore employed to map the distributed electric sources within the brain to scalp EEG recordings [Babiloni et al., 2009]:

$$\mathbf{h}_N(\mathbf{x_N}(t)) = \mathbf{L}\rho(t). \tag{5.5}$$

The observation function for fMRI is a static nonlinear function of the cerebral blood volume and the concentration of deoxyhemoglobin directly [Friston et al., 2000]:

$$h_H(\mathbf{x_H}(t)) = V_0[k_1(1-q(t)) + k_2\left(1 - \frac{q(t)}{\nu(t)}\right) + k_3(1-\nu(t))]. \tag{5.6}$$

The factors k_1, k_2 and k_3 are dimensionless but depend on the characteristics of the fMRI recording system. For 1.5 T and TE of 40 msec, $k_1 \cong 7E_0$ $k_2 \cong 2$ $k_3 \cong 2E_0 - 0.2$. $V_0 = 0.02$ is the resting blood volume fraction.

5.2.2 Neurovascular coupling

To link the two main components of the biophysical model, the neural mass model and the Balloon model, we specified three different biologically plausible neurovascular coupling mechanisms based on previous empirical results. These mechanisms are described below:

5.2.2.1 Synaptic input model

The first model assumes that the input to the Balloon model, $z(t)$, depends on the amount of nitric oxide (NO) released by synaptic activity, as originally proposed by Riera et al. [2006]. We refer to this model as the *synaptic input model*.

NO is a potent vasoactive and rapidly diffusing gas [Marshall et al., 1988], being a good candidate for regulating blood flow during functional activation [Li and Iadecola, 1994; Wang et al., 2005]. Although its synthesis is not yet fully understood, neuronal NO is thought to be generated pre-synaptically [Estrada and DeFelipe, 1998] and increases in NO concentration have been reported following increases in synaptic activity [Buerk et al., 2003].

As an aside, we note that, although in this coupling model NO is responsible for generating BOLD, the same model could be used with other pre-synaptically released substances, such as glutamate, that are thought to be responsible for increases in BOLD.

The total concentration of NO in the cortical unit is modelled as a nonlinear function, g, of the transmembrane capacitive currents in the somas of the interneurons and of the pyramidal cell. Although the genesis of NO is thought to be pre-synaptic, Riera et al. [2006] assume a direct causal relation between pre-synaptic activity and changes in post-synaptic transmembrane currents. These currents can be obtained from the derivative of the membrane potential, $I = CdV/dt$, (see Eq. 5.1) and therefore the total concentration of NO is given by:

$$C_{NO}(t) = \sum_{i=\{T,F\}} \chi_{IN} g_{IN}(C_m^0 \frac{dV^{IN^i}(t)}{dt}) + \chi_{PC} g_{PC}(C_m \frac{dV^{PC}(t)}{dt}). \tag{5.7}$$

The energetic factors χ_{IN} and χ_{PC} are introduced in order to make a distinction between relative metabolic demand in neurons of different types. C_m^0 and C_m are the effective membrane capacitances in the somas of the neurons. To take into account both inward and outward ionic currents, the nonlinear function, g, is required to be symmetric around zero and to include a saturation effect (Figure 5.1(b)):

$$\begin{aligned} g_{PC}(x) &= \rho_{PC}(1 - \exp(-x^2/w_{PC})) \\ g_{IN}(x) &= \rho_{IN}(1 - \exp(-x^2/w_{IN})), \end{aligned} \tag{5.8}$$

where $\rho_{PC,IN}$ are parameters to be estimated from the data. $w_{PC,IN}$ are fixed parameters (see Table 2 in the Appendix).

The amount of NO released in the cortical unit (Eq. 5.7) is then passed through a low-pass filter with gain A, cut-off frequency ω_0 and damping factor δ. Finally, the input to the extended Balloon model, z_{in} (and derivative r), is given by:

$$\begin{aligned} \frac{dr(t)}{dt} &= -2\delta\omega_0 r(t) - \omega_0^2 z_{in}(t) + \omega_0^2 A C_{NO}(t) \\ \frac{dz_{in}(t)}{dt} &= r(t). \end{aligned} \tag{5.9}$$

The baseline concentration of NO before stimulation, $z_0 = z_{in}(t = 0)$, is estimated from the data. In total, this model has seven free parameters (five from the hemodynamic model, $\theta_{_H}$, $\theta_{_{in}} = \{\pi_{in}, z_0, \theta_{_H}\}$, where $\pi_{_{in}} = A\chi_{_{IN}}\rho_{_{IN}}\omega_0^2$, which are estimated from the data (Table 5.1). The time series of z_{in} (input to the Balloon model) can be found in Figure 5.2 for most frequencies.

5.2.2.2 Spiking output model

For the second neurovascular coupling hypothesis we consider blood flow to be driven by the output spikes of the cortical unit, i.e the firing rate of the pyramidal cells. We refer to this model as the *spiking output model*.

The spiking activity of the layer V pyramidal cells is the outcome of information processing in the cortical unit and contains the information that is transmitted to other areas within and outside the cortex. Therefore this model looks at how BOLD signals are related to the output of local neuronal information processing as opposed to the synaptic input assessed by the previous model.

In this model the generalised logistic function (Eq. 5.2) is employed to transform the average membrane potential of the pyramidal cell population, $V(t)$, into the average rate of action potentials fired by these neurons [Wilson and Cowan, 1972] (Figure 5.1(c)):

$$z_{out}(t) = f(V(t)). \tag{5.10}$$

This model has seven free parameters (the same number of parameters of the input model), $\theta_{_{out}} = \{V_0^{PC}, \gamma_{_{PC}}, \theta_{_H}\}$, which are estimated from the data (Table 5.1). The time series of z_{out} (input to the Balloon model) can be found in Figure 5.2 for most frequencies.

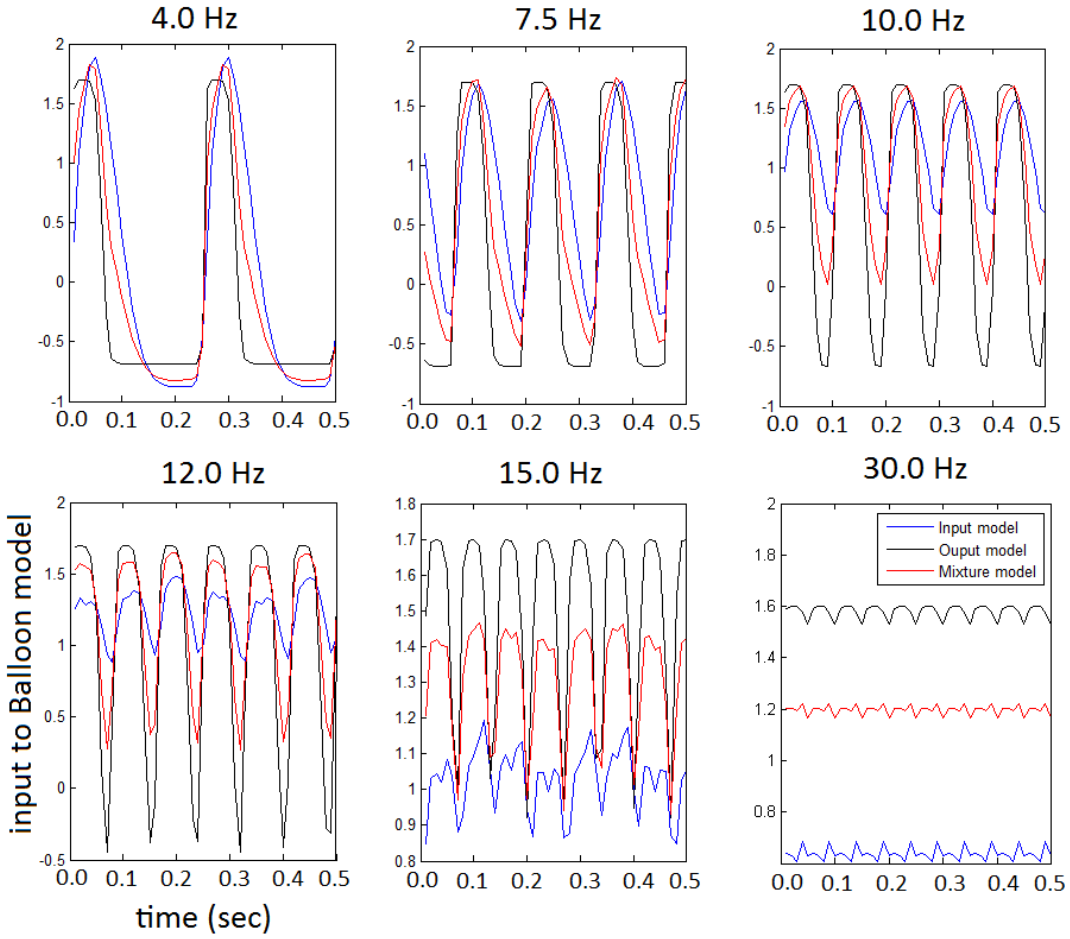

Figure 5.2: Input to Balloon model for different frequencies. Synaptic input model (blue), z_{in}, spiking output model (black), z_{out}, and mixture model (red), z_{mix}. The signals have been standardised (mean corrected and divided by the standard deviation of the signal).

5.2.2.3 Mixture model

The third coupling model assumes that both synaptic and spiking activities contribute to the generation of hemodynamic signals. Therefore, the mixture model is a sum of the amount of NO released by synaptic activity in the cortical unit and the firing rate of its pyramidal cells:

$$z_{mix}(t) = \omega_{in} z_{in}(t) + \omega_{out} z_{out}(t), \tag{5.11}$$

where ω_{in} and $\omega_{out} = 1 - \omega_{in}$ are coefficients to be estimated from the data and represent the relative contribution of each type of activity. This model has ten free parameters (three more parameters than the previous models), $\theta_{_{mix}} = \{\pi_{in}, z_0, V_0^{PC}, \gamma_{_{PC}}, \omega_{in}, \theta_{_H}\}$. The time series of z_{mix} (input to the Balloon model) can be found in Figure 5.2 for most frequencies.

5.2.3 EEG-fMRI data

5.2.3.1 Subjects and task

We use EEG and fMRI data from the same study used in Chapter 3 (see Chapter 3 for full details) to compare the neurovascular coupling models. In brief, the data were concurrently acquired for three healthy volunteers exposed to visual flicker stimuli (reversing checkerboard) of varying frequencies. Three consecutive sessions of the same experimental task were recorded for each subject. The reversing frequencies used were 4.0, 7.5, 10.0, 12.0, 15.0, 20.0 and 30.0 Hz[1]. Stimuli were delivered in epochs of 5 scans, followed by periods of 5 scans of rest (blank screen), and the order of stimulus blocks was randomised. Subjects were instructed to view a fixation cross which was visible during both rest and stimulus periods, and no overt response was required in either condition. The paradigm used here was designed to induce a large response in sensory cortex, in order to study a basic physiological mechanism, the neurovascular coupling. Although luminance levels were not held constant for the different flicker frequencies, these values were measured and taken into account by scaling the input to the model appropriately.

[1]Reversing frequencies correspond to double the value of stimulus frequencies used in Chapter 3.

5.2.3.2 fMRI data

The fMRI data were acquired and pre-processed as described in Chapter 3. In Chapter 3, we identified the brain regions activated by the flickering checkerboard in each subject. These regions are located in the subjects' visual cortex, as expected (see Figure 5.3(a) for an example subject). The coordinates of the corresponding cluster maxima are: $[27, -71, -9]$ mm, $[18, -104, 21]$ mm and $[-9, -101, 12]$ mm (Talairach coordinates) for each subject, respectively. From these location we extracted the BOLD signal (200 scans per session) by calculating the first principal component of the adjusted data (removing the global drift and other confounds) from voxels within a 6 mm spherical volume centered on the cluster maximum. The resulting time-series for each session were then epoched and averaged (in the time domain) across epochs (Figure 5.3(b)). These time-series were used to estimate the parameters of the neurovascular coupling model, as described below.

5.2.3.3 EEG data

The EEG data were also acquired and pre-processed as described in Chapter 3. Here we use the scalp steady state visual evoked responses (SSVERs) to reconstruct the electrical activity at the source level. SSVERs were computed by first epoching the artefact-corrected 27-electrode EEG data acquired inside the MRI scanner, for each session, in a 15-second post-stimulus window and then averaging (in the time domain) across trials. This procedure yielded 7 averaged 15-second time-series for each session corresponding to the 7 different flicker frequencies used. The source electrical activity was then obtained as follows. Given a source region with known anatomical location, we can form the $N_C \times 1$ lead field vector $\mathbf{L}$ where N_C is the number of EEG sensors. This vector was obtained with SPM8 using a template mesh for the location and orientation of the cortical source and a boundary element method for the head model. The source location was chosen to be the corresponding cluster maximum identified with the fMRI data (see previous section). Given that the number of sources ($N_S = 1$) is smaller than the number of scalp channels ($N_C = 27$), activity in the source region can be estimated as follows [Baillet et al., 2001a]:

$$\rho^k(t) = \mathbf{L}^+ \mathbf{y}_N^k(t), \tag{5.12}$$

where $\mathbf{L}^+$ denotes the Moore-Penrose pseudo-inverse of the lead field vector $\mathbf{L}^+ =$

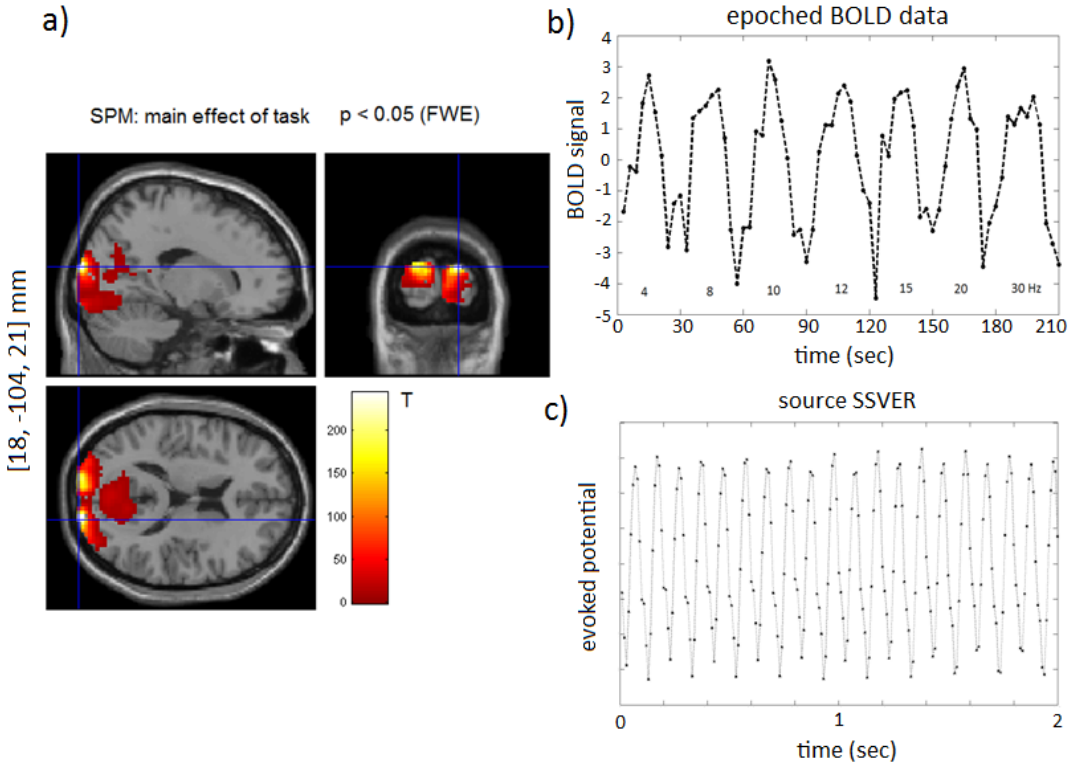

Figure 5.3: EEG-fMRI data. a) SPM results (3 sessions, example subject): effect of visual flicker stimulation on fMRI data. The voxel location corresponds to the most significant cluster maximum (Talairach space), p-value < 0.05 (FWE). b) Epoched BOLD signal (eigenvariate) from the most significant cluster maximum - one example session. c) 2 second source SSVER, ρ, from the same cluster peak from 1 example session and frequency (10 Hz). Both signals have been standardised (mean corrected and divided by the standard deviation of the signal).

$(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T$. Here $\mathbf{y}_N^k(t)$ is the artifact-free SSVER for frequency k and one session. The resulting source time-series (for all frequencies and all sessions), $\rho(t)$, were used to estimate the parameters of the neural mass model (see below) (Figure 5.3(c)). This renders observation equation 5.5 redundant as we have already projected our EEG data into source space.

5.2.4 Bayesian model inversion and selection

Using EEG-fMRI data in combination with Bayesian inference allows us to estimate the underlying synaptic and spiking activities, along with other parameters of the biophysical framework. Additionally, we can compare the different neurovascular coupling hypotheses using Bayesian model evidence.

As described in Chapter 2, the posterior density used in Bayesian inference is an optimal combination of prior knowledge and new observations, and provides a complete description of uncertainty about the parameters. Here we use priors based on empirical knowledge for both the neural mass model parameters and the coupling/hemodynamic parameters. These priors correspond to the parameter estimates obtained by Riera et al. [2006]. Prior variances were chosen to be of the same order of magnitude as the prior means to ensure a coefficient of variance ($CV = \frac{\mu}{\sigma}$) of approximately 1 for all parameters.

The coupling models were inverted as described in Chapter 2, Section 2.2, assuming the error covariance is isotropic over the EEG and fMRI predictions $C_\epsilon^N = \sigma_N^2 I$ and $C_\epsilon^B = \sigma_B^2 I$.

Due to the fact that we are using both EEG and fMRI datasets to invert the models, which have very different sampling periods (0.01s and 3s, respectively), we developed an efficient multi-step inversion approach, described below.

After inversion, Bayes factors, Eq. 2.20, were used to compare the models, as introduced in Chapter 2.

5.2.4.1 Multi-step inversion

The use of both EEG and fMRI data to estimate the electro-vascular model is affected by the difficult problem of how to deal with the disparity between the two datasets' time scales. In our study, for each fMRI point (sampled every 3 secs) we have 300 EEG data points (sampled at 100 Hz). The large amount of EEG data renders the

model inversion computationally intensive, as for each parameter update we must integrate the model equations at a fine temporal scale (1000 Hz).

To overcome this problem we developed a computationally efficient inversion scheme based on partitioning model inversion into separate steps depending on the time-scales of the data involved. We refer to this scheme as a 'multi-step inversion' approach. This procedure generalises to other datasets and can be used in other multimodal studies, such as MEG-fMRI or LFP-fMRI, where the amount of data and time scales are very different between modalities.

This 'multi-step inversion' approach works as follows (Figure 5.4):

(1) First we selected 2 secs of the source SSVERs (Eq. 5.12) for each frequency (4 to 30 Hz) and session to identify the electrical states, $\mathbf{x}_N$, and parameters, θ_N of the NMM. Using the EEG data alone to estimate the parameters of the NMM makes sense because these data are not dependent on the changes in the vasculature that give rise to BOLD. We chose to fit only 2 secs for each frequency (concatenated and chosen from the middle of the stimulation block to avoid onset and offset transients) because, as reported in Riera et al. [2007], the averaged signal for the entire 15 secs is very regular (stationary), being sufficient in our view to estimate the model without using the entire trial block. Reducing the data to 2 secs per frequency considerably speeds up the inversion process. However, it is worth noting that by assuming stationarity of the response over 15 secs and only estimating the model from 2 secs, one can potentially be losing important slower variance components that might be informative/important for looking at EEG-fMRI coupling. The parameters for each session were estimated iteratively using a time step of 1 msec. At each iteration the predictions were downsampled by a factor of 10 in order to fit the 100 Hz source SSVER data. Here we assume the neuronal response is stationary within a given epoch (15 sec stimulus interval) with averaged EEG and BOLD signals used here.

(2) After estimating the electrical parameters (previous step), we used these estimates to integrate the full LEV model. Importantly, this integration takes place only once (as opposed to a 'single-step' approach, where it would have to be integrated at every iteration). The integration is implemented as above but instead of 2 secs, the input to the model is now 15 secs of stimulation and 15 secs of rest for each frequency. We integrate the full models with the three different coupling mechanisms described above and produced the following time-series as our input to the BOLD response (next step). For the synaptic input model the output time-series is the total

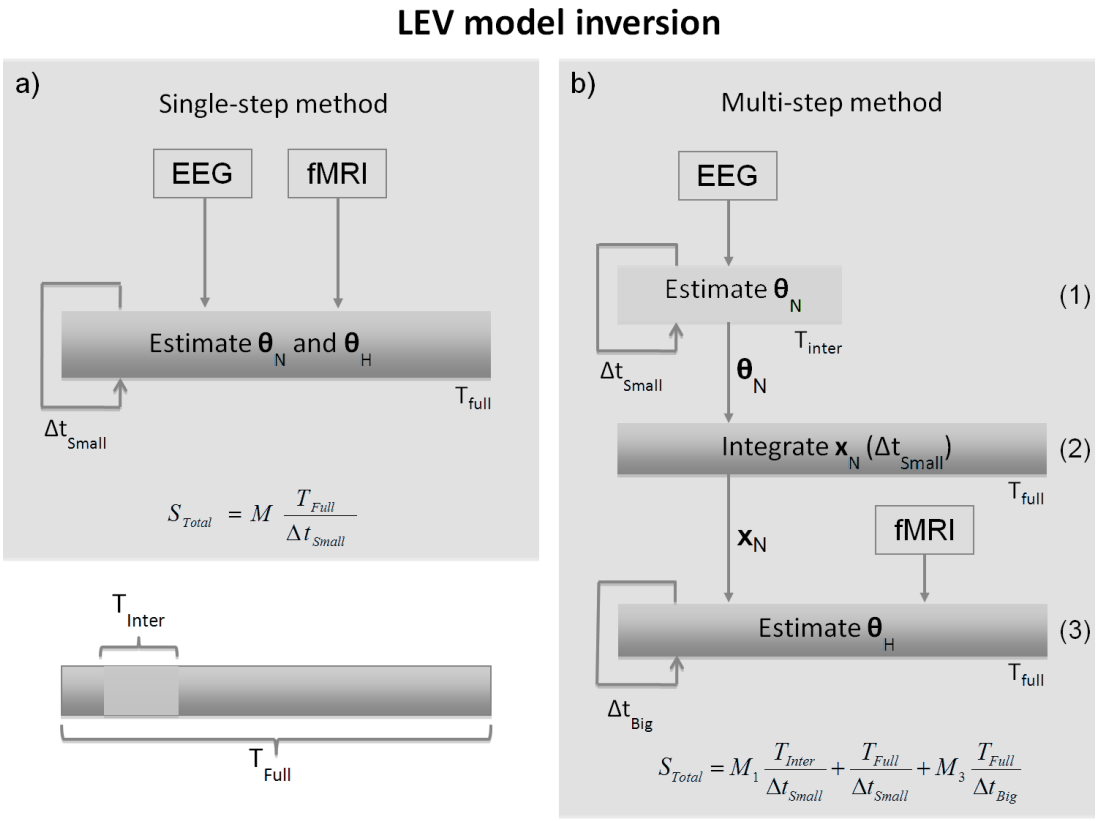

Figure 5.4: LEV model inversion: here we adopted a 'multi-step' approach as opposed to inverting the model in a single step. a) Single-step approach: the EEG and fMRI data are used to estimate the neuronal and hemodynamic parameters (θ_N and θ_H) simultaneously. At each iteration the model equations are integrated at a small time scale matching that of neuronal activity, Δt_{small}, for the entire time interval, T_{full} (15 seconds for each frequency). b) Multi-step method: here the inversion is performed in three main steps. (1) First the neuronal parameters, θ_N, are estimated (using M_1 iterations) from the EEG data with a fine temporal resolution, Δt_{small}, but for a smaller period, T_{inter} (2 seconds for each frequency). (2) In the second step these parameter estimates are used to integrate the neuronal equations of the LEV model, $\mathbf{x}_N$, with the same temporal resolution Δt_{small} but entire time interval T_{full}. (3) In the last step we use the BOLD data to estimate (using M_3 iterations) only the hemodynamic parameters, θ_H, with a lower time resolution of Δt_{big} over the full time interval, T_{full}. The total number of time steps, S_{total}, for each approach is displayed in each gray box.

NO concentration, Eq. 5.9. For the spiking output model the output time-series is the firing rate of pyramidal cells, Eq. 5.10, whilst for the mixture model both of these output time-series were produced, Eq. 5.11. These output time-series were downsampled to 10 Hz to reduce the estimation time of the next step and used as inputs to the Balloon model.

(3) Finally, with the time-series for all coupling models obtained in the previous step we estimated the extended Balloon model using the epoched BOLD data for all frequencies. The estimation was again performed iteratively as described above, this time with a 100 msec time step because the vascular dynamics is a much slower process than the electrical processes. The value of the free energy (surrogate to the log model evidence) for each neurovascular model was then used to infer the optimal coupling mechanism.

5.3 Results

5.3.1 Synthetic data

In this section, simulations are used to explore the behaviour of the model and its ability to reproduce EEG and BOLD data under the experimental conditions described in the previous section. The responses of the three neurovascular coupling models to changes in stimulus frequency are also shown. These synthetic signals are used to test the model inversion routines and to verify that Bayesian model comparison can be used to infer the correct coupling model.

The LEV model was numerically integrated using the multi-step Adams-Bashforth-Moulton predictor-corrector algorithm implemented in the MATLAB (The MathWorks, Inc.) function *ode113*. The integration step used was 1 msec (1000 Hz) for the electrical and vascular states. The integrated signals were then downsampled to 100 Hz in the EEG case and to 0.3 Hz for the BOLD signal. The input to the model is described below.

5.3.1.1 Model input

The input to the LEV model was generated by creating a series of single events with the same frequency as the reversing checkerboard (4.0, 7.5, 10.0 ... Hz). These events are modelled as Gaussian functions of $\sigma = 17.0$ msec width: $I^+(t) = \sum_i A \exp(-|t -$

$t_i|^2/2\sigma^2$). This value of σ corresponds to the screen refresh interval. The amplitudes A are fixed over time but differ for excitatory versus inhibitory populations. In our simulations we used the amplitudes $A_1 = 1$pA and $A_1 = 0.4$pA for I_1^+ and I_3^+, respectively, as proposed in Riera et al. [2006]. These amplitudes are estimated from the data when using the EEG-fMRI signals (see below). Input I_3^+ was also delayed by 100 msec with respect to I_1^+ as suggested in Riera et al. [2006]. Cortico-cortical interactions were neglected and so I_2^+ was set to zero during the entire period of integration. Due to the fact that luminance levels were not kept constant for the different frequencies we multiplied the input time-series according to the lux measures (from low to high frequencies) by: 1.00, 0.96, 0.93, 0.91, 0.88, 0.82, 0.74 (lower frequencies had higher luminance levels).

5.3.1.2 Frequency-response curves

We first generated data from the LEV model separately for the different stimulus frequencies (4 to 30 Hz). We used the three neurovascular coupling mechanisms described above. The data were simulated using the parameter values summarised in Table 1 and Table 2 (Appendix) for a period of 15 seconds of stimulation and 15 seconds of rest. The simulated signals showed that all coupling models predict an increase of the BOLD signal during stimulation, as expected, and synchronisation of the EEG signal to the input frequency. Figure 5.5 shows the EEG and fMRI signals generated for a period of 15 sec of stimulation and 15 sec of rest using the synaptic input model.

We then looked at the behaviour of the fMRI signal predicted by the different coupling models for all frequencies. Figure 5.6 presents the frequency-response curves obtained. These curves correspond to the maximum amplitude of the BOLD signal for each stimulus frequency. As can be seen in Figure 5.6, the synaptic input model predicts an increase in the BOLD response until approximately 8 Hz and a decrease afterward. This result confirms the simulations of Riera et al. [2006] who found a similar frequency-response curve for the NO mechanism between (0.5 and 16 Hz). In addition, this result validates the use of a deterministic model instead of the original stochastic model. The stochastic effects are therefore not necessary to reproduce the frequency response curve obtained in Riera et al. [2006]. Contrary to the synaptic input model, the spiking output model predicts an increase in the BOLD response with input frequency without any saturation effect (Figure 5.6).

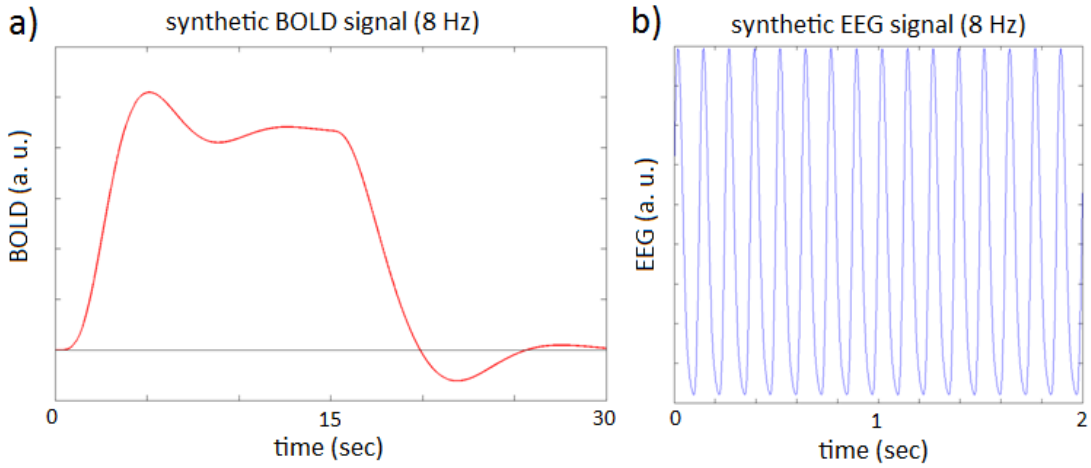

Figure 5.5: Simulated data: a) BOLD response for a stimulation block (15 seconds of stimulation and 15 second of rest) of 8 Hz reversing frequency; b) EEG signal for the same stimulus (2 seconds). Both signals have been standardised (mean corrected and divided by the standard deviation of the signal) as used for model inversion.

Figure 5.7(a) shows the frequency-response curve for the real fMRI data. For real data the values plotted in this curve correspond to GLM coefficients as a function of frequency (stimulus). These are obtained when we regress the BOLD signal using the onsets of the stimuli as our regressors, or columns of the design matrix. Each column corresponds to a different frequency and the associated coefficient tells us how much BOLD is expected to increase with that particular frequency. As can be seen in Figure 5.7(a), the response of the real BOLD signal to the different frequencies also peaks at 8 Hz and has a minimum at 15 Hz. This behaviour has been previously reported in human BOLD data for frequencies below 16 Hz under similar experimental conditions [Parkes et al., 2004; Singh et al., 2003; Wan et al., 2006]. Above 15 Hz this curve has a second peak in BOLD signal amplitude at 20 Hz and a decrease afterward (Figure 5.7(a)). The same type of curve is reported in Parkes et al. [2004]: two maxima at 8 and 20 Hz, a smaller peak at 12 Hz, and the rest of the frequencies ($\leq$ 20Hz in Parkes et al. [2004]) lie below these values.

The frequency-response curve for the measured SSVERs is plotted in Figure 5.7(b). The curves for all three sessions of an example subject show a peak at 12 Hz and a decrease in amplitude afterward. This same curve was found in all other subjects and sessions. This means the peaks in the BOLD signal cannot be explained from the electrical signals alone.

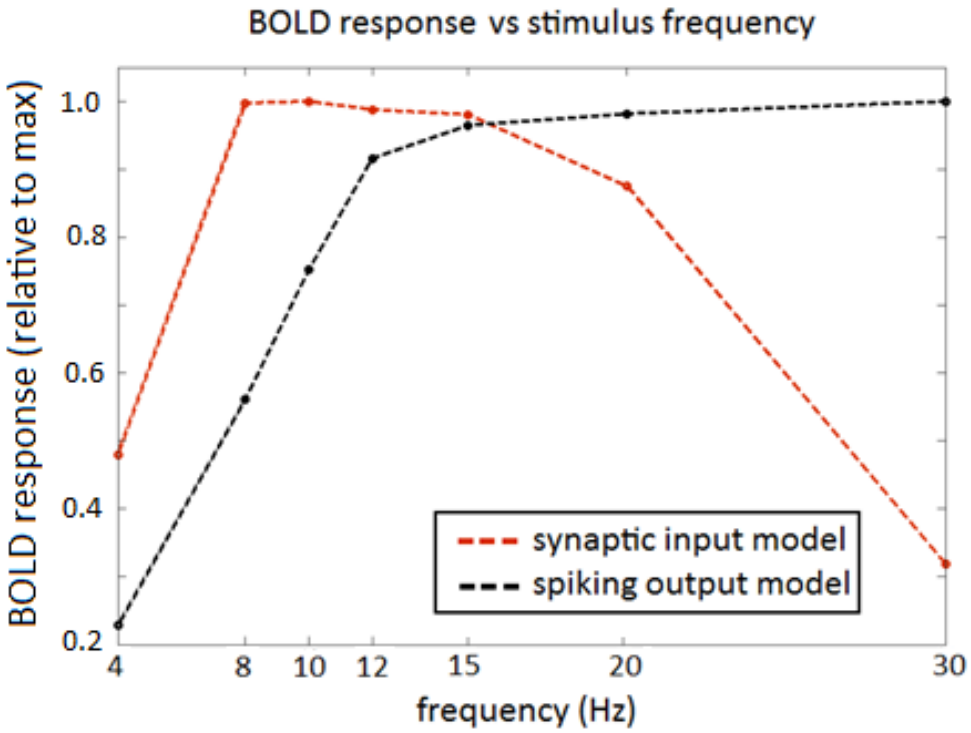

Figure 5.6: Model frequency response curves -synthetic data: a) Predicted BOLD response versus reversing frequency for the synaptic input and spiking output models. The curves show the BOLD response obtained for each stimulus frequency (divided by the maximum peak for each model).

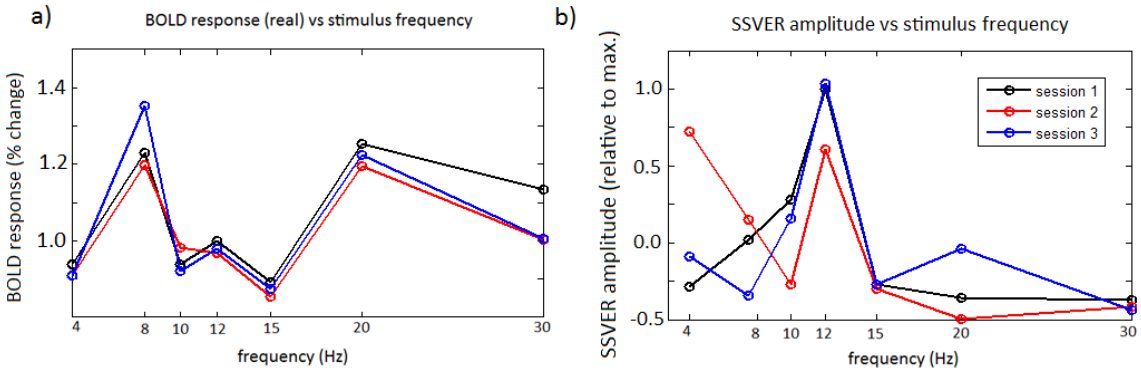

Figure 5.7: Measured frequency response curves - EEG-fMRI data: a) Measured BOLD response versus reversing frequency. The values on the y-axis correspond to per cent changes of the global mean signal. b) Frequency-response curve for EEG data. Each point corresponds to the amplitude of the evoked response (divided by the maximum response) at that frequency ($\sum_m |\text{SSVER}_m|^2$), where m is the number of bins. The maximum value was $2.07\mu\text{V}^2$.

5.3.1.3 Model parameters

Table 5.1 lists the parameters for the electric, $\theta_{_N}$ and vascular, $\theta_{_H}$, components of the model that are estimated from the (synthetic) data. These are the same parameters estimated in Riera et al. [2007]. We also summarise the coupling parameters in the same table: $\theta_{_{in}}$, $\theta_{_{out}}$ and $\theta_{_{mix}}$ (Table 5.1). The amplitudes of the three input currents (I_1^+, I_2^+ and I_3^+) and $\theta_{_N}$ are estimated from EEG in step (1) of the inversion. $\theta_{_H}$ are estimated from the BOLD signal in step (3). $\theta_{_{in}}$ and $\theta_{_{out}}$ and $\theta_{_{mix}}$ are estimated from both EEG and fMRI data in steps (1) and (3) of model inversion.

When using the observed EEG and fMRI signals, the priors on the parameters corresponded to the parameter estimates obtained by Riera et al. [2007], that is, from the inversion of the same electro-vascular model with similar EEG-fMRI data.

5.3.1.4 Model comparison

We then tested if Bayesian model comparison could be used to correctly decide upon which coupling model was used to generate the data, and if despite the small number of samples of fMRI compared to EEG we could still infer the right model.

We again generated data using the three coupling models as described above. We generated data for all the frequencies concatenated, with additive Gaussian observation noise: $\sigma_{_N} \sim N(0, 0.3I)$ and $\sigma_{_B} \sim N(0, 0.008I)$. These values are based on the

Table 5.1: Estimated parameters: these are the parameters estimated from synthetic (column 7) and measured EEG-fMRI data (last column). These are from one example session and for all frequencies. The parameter π_{in} for the synaptic input model corresponds to: $\pi_{in} = A\chi_{IN}\rho_{IN}\omega_0^2$ (see Eq. 13-18 and Table 2 in Appendix A). 'Symb.' means Symbol, 'Estim.' means Estimated, 'Obser.' means Observed and 'Mix.' means Mixture.

Electrical, vascular and coupling parameters							
					Synthetic		Obser.
Type	Description	Symb.	Units	Prior	True	Estim.	Estim.
θ_N							
	Synaptic input	I_1^+	pA	1.00	0.80	0.85	0.94
	Synaptic input	I_2^+	pA	1.00	1.00	1.00	1.00
	Synaptic input	I_3^+	pA	1.00	0.50	0.60	0.60
	GABAergic IN synaptic factor	α_{IN}	pA	0.30	0.50	0.49	0.53
	PC voltage-ampere (VA) function	V_0^{PC}	mV	0.60	0.90	0.78	0.42
	PC VA function	γ_{PC}	mV^{-1}	6.00	4.00	5.62	5.95
θ_H							
	Signal decay	τ_s	ms	0.65	0.50	0.65	0.59
	Autoregulation	τ_f	ms	0.41	0.28	0.41	0.40
	Transit time	τ_0	ms	0.98	0.78	0.98	0.91
	Stiffness	α	no dim.	0.32	0.25	0.32	0.32
	Resting O2 extraction fraction	E_0	no dim.	0.34	0.30	0.34	0.34
Coupling							
NO (θ_{in})	NO concentration baseline	z_0	no dim.	0.10	0.30	0.29	0.29
	NO synaptic current factor (IN)	π_{in}	s^{-2}	1590	1500	1590	1590
FR (θ_{out})	PC VA function	V_0^{PC}	mV	0.78	0.90	0.63	0.17
	PC VA function	γ_{PC}	mV^{-1}	5.62	4.00	5.70	7.98
Mix. (θ_{mix})	NO coefficient	ω_{in}	no dim.	0.50	0.40	0.40	0.29
	FR coefficient	ω_{out}	no dim.	0.50	0.60	0.60	0.71

signal-to-noise ratio for the observed data (1 for the averaged EEG signals and 2 for the averaged BOLD signals). We then fitted the coupling models to each of the three synthetic datasets.

We verified that Bayesian model comparison inferred the correct model in all cases, with a minimum Bayes factor of approximately 20 (log-Bayes factor of 3) (Figure 5.8). This value corresponds to strong evidence in favour of the model that generated the data and a posterior model probability over 0.95 [Penny et al., 2004]. The parameter values used to generate the data and the corresponding parameter estimates and priors for each model can be found in Table 5.1. As can be seen, the parameter estimates were close to the true values used in data generation, except for V_0^{PC}. For this parameter the estimated value was found to be farther away from the true value than the prior. This result might be due to a bad choice of true or prior value and potential identifiability issues in the estimation of V_0^{PC} and γ_{PC}. Despite this fact, the fit of the output model was found to be good (not shown).

As an aside, we note that, as with any gradient-ascent based optimisation algorithms, our inversion scheme is subjected to the possibility of running into local minima. One way to tackle this problem can be to initialise the inversion in different parameter regimes. In this work we have only observed once a clear case of local minimum, where the fit of one of the models to one session was extremely poor. We have then initialised the parameters with the estimates from other sessions and the inversion scheme was able to find new parameter estimates that provided a good fit to the data, similar to what was obtained for the other sessions.

5.3.2 EEG-fMRI data

Finally we fit the electro-vascular model with the three different coupling mechanisms to the EEG and fMRI data. We used the same 'multi-step' inversion procedure described in the previous section. Figure 5.9 shows the model predictions for EEG, as well as predictions of the coupling models and the BOLD response.

5.3.2.1 Model comparison

Our analysis focused on the relevant contributions of synaptic and spiking activity models as a function of stimulation frequency. To this end we divided the stimuli into 'low-frequencies' (4 to 15 Hz), 'high-frequencies' (10 to 30 Hz) and 'all-frequencies' (4 to 30 Hz) and the analysis was repeated for these three regimes. A summary of the

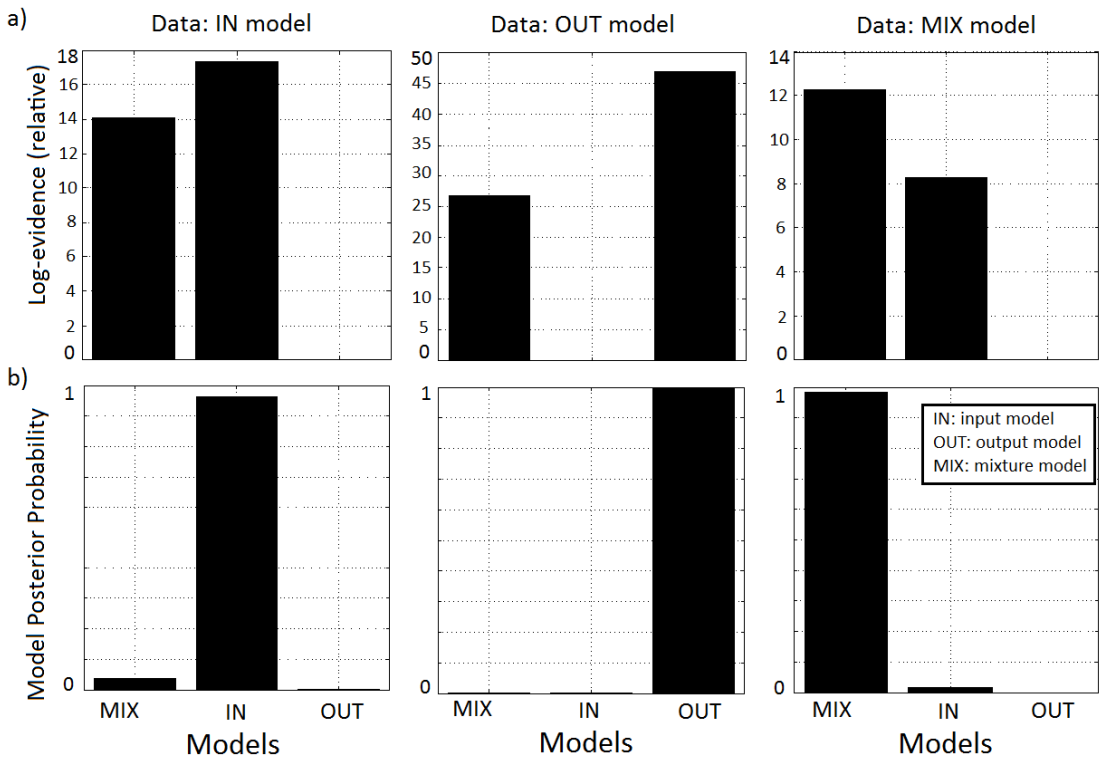

Figure 5.8: Model comparison with synthetic data: we generated data with the different coupling models (IN: synaptic input model; OUT: spiking output model; MIX: mixture model). We then fitted these datasets with the same three coupling models and obtained the results plotted in the figure. a) Difference in log-evidences relative to worst model. b) Corresponding model posterior probabilities.

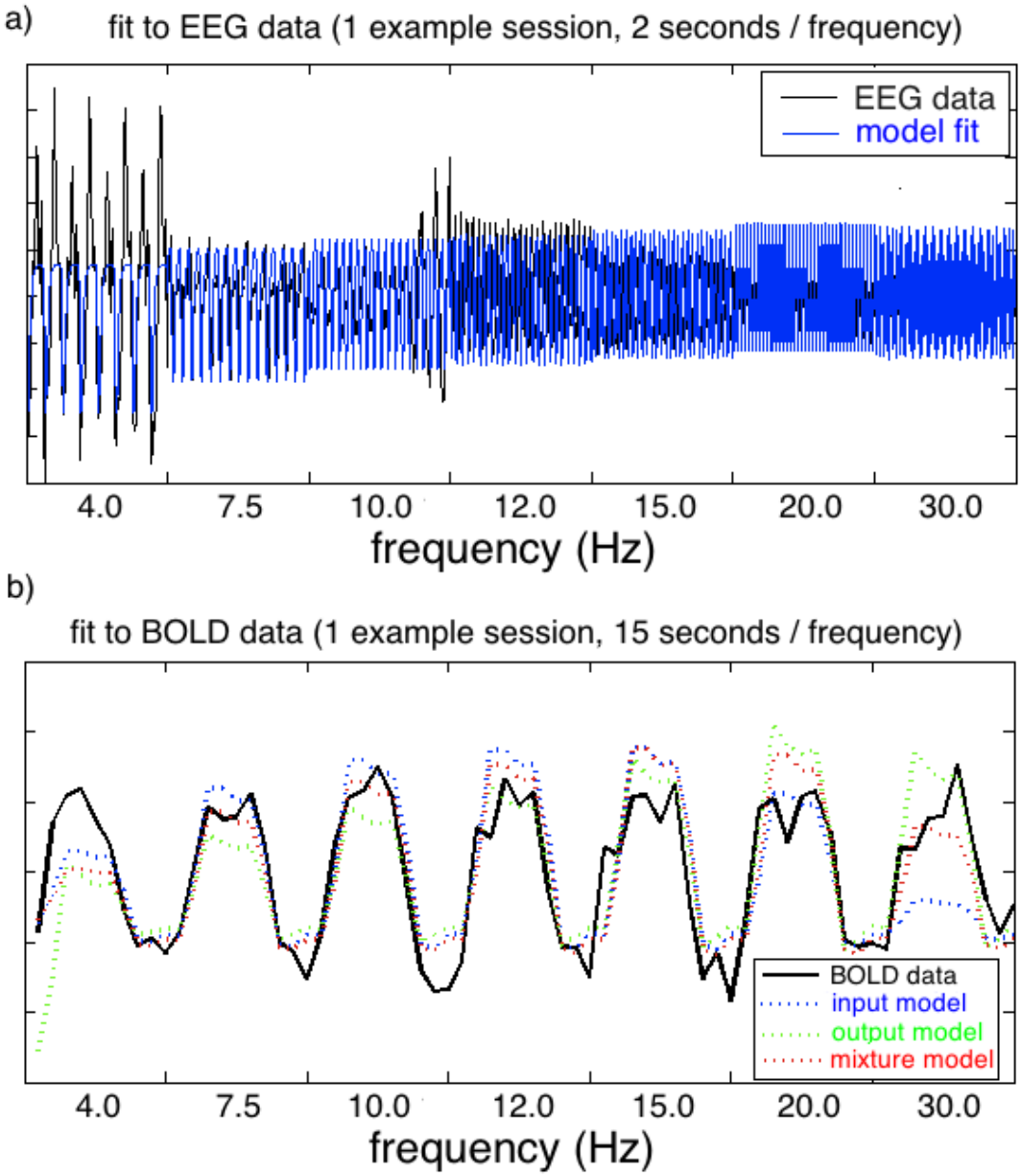

Figure 5.9: Model identification: a) EEG time-series (black line) and model fit (blue line) for one example session and subject (2 seconds of data per frequency). b) BOLD time-series (solid black line) and model fit (dotted lines) for one example session and subject (15 seconds of data per frequency). The M-shaped response of Fig. 5.7 cannot be observed here because the BOLD signal is confounded by luminance levels, which are accounted for in the regression analysis that was done to obtain the frequency-response curves of Fig. 5.7. For this example session, the model evidence is greater for the Mixture model than for the other two models. This can be seen in the figure, i.e. the mixture model (red) provides the best overall fit to the data of this session. The signals have been standardised (mean centred and divided by the standard deviation of the signal) as used in the model inversion scheme.

model comparison results for all subjects can be found in Figure 5.10. The results for all sessions, subjects and frequency regimes can be found in Table 3 of the Appendix.

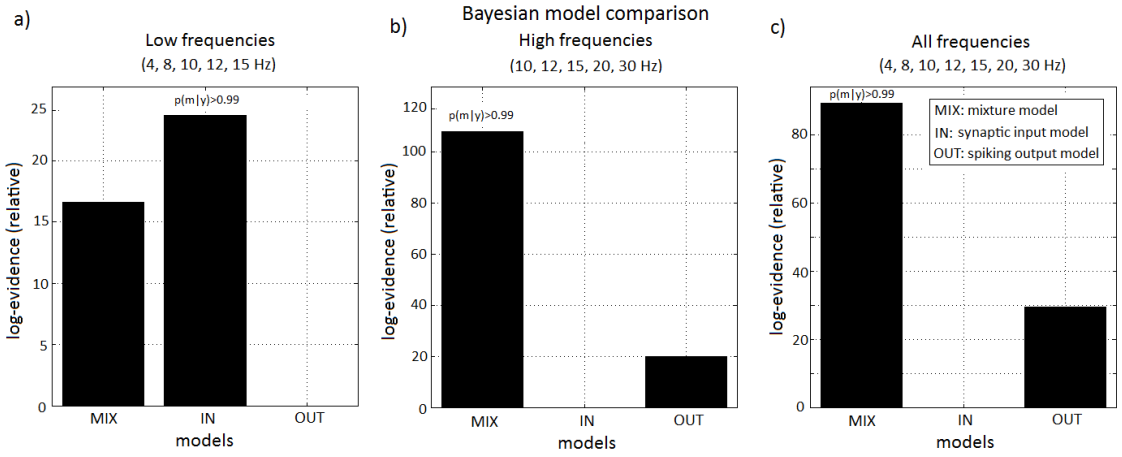

Figure 5.10: Model comparison (MIX: mixture model; IN: synaptic input model; OUT: spiking output model): log-model evidence relative to worst model (for low, high and all frequencies). These are group results for all subjects and sessions analysed (the log-evidences are summed over subjects).

As can be seen in Figure 5.10(a), in the low-frequency regime we found that the synaptic input model best explained the observed data. In this regime the spiking output model was the worst model. The difference in log-model evidence between the best model (synaptic input) and the second best model (mixture) was above 5. This value corresponds to strong evidence in favour of the synaptic model and a probability, p, over 0.99 of this model being the best model to explain the data in this regime [Penny et al., 2004]. This result was consistent accross subjects and sessions analysed (5 out of 6 sessions) (Table 3 of the Appendix). The model evidence values can be found in Table 3 of the Appendix.

However, when we analysed the high frequencies, the mixture model was found to be the best model with probability $p > 0.99$ (Figure 5.10(b)). This result was again consistent across subjects and for the majority of sessions (7 out of 9 sessions) (Table 3 of the Appendix). In this regime the spiking output model was the second best and, contrary to the low-frequencies case, synaptic activity contributed the least to the BOLD response.

For both regimes, the inferred neuronal firing rates were found to be commensurate with the stimulation frequency. Finally, an additional analysis across all frequencies

revealed that the mixture model was the best model, again with probability $p > 0.99$ (Figure 5.10(c)). This result was found in 6 of the 9 sessions analysed, although in one of the sessions the model evidence for all three models was nearly identical (Table 3 of the Appendix).

With hindsight it seems unsurprising that when we analyse all frequencies the mixture model was found to explain the data better than the input and output models alone. As we observe in Figure 5.7(a), the double peaked frequency-curve of fMRI data can be easily explained by a weighted combination of the frequency-response curves predicted for the input and output models individually (Figure 5.6). This weighted combination is the definition of the mixture model and the weights (mixture parameters) depend on the regime of frequencies analysed, providing, for instance, a one-peaked or two-peaked curve for low and all-frequencies, respectively.

These results were robust to the choice of partition into low/high frequencies. Similar results (not shown) were obtained with partitions such as: low-frequencies (4, 8, 10, 12 Hz) and high-frequencies (15, 20, 30 Hz).

5.4 Discussion

In this chapter we used EEG-fMRI data and a biophysically informed mathematical model to investigate the relationship between neuronal activity and the BOLD signal in human visual cortex. In particular, we explored the contributions of synaptic input and spiking output activities to the generation of the BOLD response.

We have provided evidence that the BOLD signal is dependent upon both synaptic and spiking activity but that the relative contribution of these two factors are dependent upon the underlying neuronal firing rate, which is proportional to the stimulation frequency. When the underlying neuronal firing is low then BOLD signals are best explained by synaptic input, in agreement with previous animal studies, such as Logothetis et al. [2001]. This result is also in line with more recent studies, such as Viswanathan and Freeman [2007] and Rauch et al. [2008], which show that the BOLD response is only affected by changes in synaptic-related activity (measured with LFPs) and not by changes in spiking activity (measured with MUA) when these two signals can be dissociated.

However, when the stimulation frequency, and correspondingly the neuronal firing rate, is high then both synaptic and spiking activity are required to explain the BOLD

signal, as observed in, for example, Heeger et al. [2000] and Mukamel et al. [2005]. We were particularly encouraged to find that a combination of synaptic input and spiking output frequency response curves (Figure 5.6) can explain the doubly-peaked BOLD response observed by Parkes et al. [2004] and replicated in our own data.

One possible explanation for the increased performance of the output model with higher frequencies comes from neuroenergetic studies such as e.g. Smith et al. [2002] and Maandag et al. [2007]. In these studies brain metabolism was found to depend strongly on neuronal spiking, with increases in oxygen consumption reflecting higher firing rates. More recently, Lee et al. [2010] have found that differences in the BOLD response between different brain areas (motor cortex and thalamus) could be explained by underlying differences in the firing rates of the corresponding neuronal populations.

Our results also support the conclusion that the relationship between synaptic activity, spikes and BOLD signals depends on the specific neuronal circuitry engaged in task processing. Moreover, one can speculate that different coupling mechanisms involving different types of cells and molecules could come into play depending on the task in question.

Despite our concern about the small number of fMRI samples compared to EEG, our initial results with synthetic data showed that it is possible to make inferences on different hypotheses for the neurovascular coupling using a generative modelling framework and Bayesian model comparison. The issue of different time-scales was addressed by partitioning the estimation of electrical and vascular states into a multi-step approach. In this approach we first estimated the electrical states and parameters from the EEG data and then integrated the full electro-vascular model using these estimates. From the integrated model we extracted the input time-series to the Balloon model, which we then inverted using BOLD data. The last two steps were repeated for each coupling model.

This method significantly increases the computational efficiency of the model inversion. However, this multi-step approach is only possible with a deterministic model. In this work we used a deterministic version of the stochastic electro-vascular model proposed by Riera et al. [2006]. Under different experimental conditions, which do not induce a large sensory response, the introduction of stochastic effects might be essential to reproduce the empirical data. In this case, other Bayesian inversion frameworks can be employed to estimate the model parameters, such as Friston et al. [2008] and Daunizeau et al. [2009].

It is also worth noting that despite the fact that the mixture model had more parameters than the input and output models, this extra complexity did not provide a significantly better fit to the data in the low-frequency analysis than the input model. This complexity is correctly penalised using Bayesian methods, such as the one used here (see Chapter 2).

One concern about the coupling models defined here regards the definition of NO concentration. As mentioned in the Methods section, NO is thought to have a pre-synaptic synthesis [Estrada and DeFelipe, 1998; Wang et al., 2005]. However, here and in Riera et al. [2006] the concentration of NO is modelled through post-synaptic quantities such as the transmembrane capacitive currents. Although in principle these two phenomena are directly related (increases in pre-synaptic activity mean larger post-synaptic effects) this is not always the case. Changes in transmembrane currents at the post-synaptic level can be caused by different processes such as chemical-gated channels, electric-gated channels, and passive leakage, not all of them being related to pre-synaptic activity. Therefore the transmembrane currents are an indirect way of quantifying the amount of NO released during synaptic activity. However, this issue is also encountered in experimental measures of synaptic activity, such as local field potentials. This signal is a surrogate post-synaptic signal, which is also affected by other slow potentials occurring at the cellular level that do not have a purely pre-synaptic origin.

To our knowledge this work presents the first quantitative model comparison of different biologically plausible mechanisms for neurovascular coupling in human cortex using EEG-fMRI data and a realistic biophysical model.

However, even though our results were consistent across the three subjects and the majority of sessions, the case study approach adopted here has its limitations. Namely, it does not quantitatively address the issue of inter-subject variability and it therefore precludes inferences at the population level. With a larger sample of subjects, inter-subject variability can be accommodated using the Random-Effects (RFX) model selection approach developed by Stephan et al. [2009] and reviewed in Chapter 2 and Chapter 4.

One possible technical improvement to our work would be to fit the EEG data in the frequency rather than in the time domain. This would be a sensible option as we have assumed that the EEG is stationary during the time periods of interest. This approach is used in related work [Moran et al., 2008], where EEG spectra are related to parameters of underlying neural mass models.

The work in this chapter has investigated models of neurovascular coupling for neurodynamic and hemodynamic activity in a single brain region. In future, we envisage this work being extended to activity in multiple brain regions. The LEV model [Riera et al., 2006] is particularly relevant here because the cortical unit is explicitly modelled as receiving separate inputs from thalamus and other cortical regions. This would allow us to investigate which aspects of BOLD are due to synaptic input from cortex versus thalamus. Further, this would bring the research into the more general area of brain connectivity modelling, which is the topic of the following chapter.

We hope that future studies with other datasets and different experimental conditions will employ our modeling approach so that a balance of evidence can be reached that clearly disambiguates between different hypotheses concerning neurovascular coupling.

Understanding the underlying biophysical mechanisms behind the coupling between neuronal activity and the BOLD response is vital not only for improving the interpretability of the BOLD response, but also for relating findings from fMRI research with results from other neuroscientific disciplines. For example, a large amount of work in neuroscience relies on recordings of firing rates from single responsive neuronal cells.

Chapter 6

Post-hoc selection of dynamic causal models

6.1 Introduction

Dynamic causal modelling (DCM) is a mathematical framework to estimate, and make inferences about, the coupling among brain areas and how this coupling is influenced by changes in experimental context [Friston and Penny, 2003]. Although it was originally introduced as a hypothesis driven procedure, in which a small number of neurobiologically motivated models are compared, recently, a trend has emerged for comparing very large numbers of models in a more exploratory manner.

Bayesian model selection (BMS) is a powerful method to compare different models for explaining observed data. BMS is based on the model evidence, which is the probability of the data given the model. Even though this quantity is not, in general, straightforward to compute, it is now well established that statistical models can be compared using a variational free energy approximation to the evidence [Beal and Ghahramani, 2003]. This approximation has widespread application, and, in neuroimaging, it has become the method of choice for comparing models of effective brain connectivity, in particular dynamic causal models (DCMs) [Penny, 2012; Stephan and Friston, 2010].

Model comparison in this context has hitherto proceeded by individually fitting all competing models to data and then approximating the model evidence with the variational free energy bound [Friston et al., 2007a]. We refer to this approximation to the model evidence as the optimised evidence.

Very recently, Friston and Penny [2011] have proposed an alternative, post-hoc, approximation to the model evidence that is computed by fitting only the very largest of a set of models: a full model from which all other (reduced) models can be formed by removing model parameters. This scheme approximates the evidence for any nested model within a larger model using only the posterior density of the full model. We refer to this approximation as the post-hoc evidence.

The benefit of this post-hoc approach is a huge reduction in the computational time required for model fitting. This is because only a single model is fitted to data. This means that a potentially huge model space can be searched relatively quickly.

In addition to the model evidence approximation, Friston and Penny [2011] also propose a way to estimate the connectivity parameters for all reduced models from the posterior density over the parameters of the full model. More specifically, the posterior mean and precision of the reduced model can also be determined solely from the mean and precision of the parameters of the full model.

The post-hoc approach can also be viewed as a generalisation of the well-known Savage-Dickey density ratio [Dickey, 1971] (see Figure 6.1), in which the reduced models have certain parameters fixed at zero. To our knowledge, the Savage-Dickey method [Dickey, 1971], has not yet been applied to neuroimaging problems, although it has been applied in other fields, from cognitive psychology [Wagenmakers et al., 2010] to cosmological models [Trotta, 2007]. The recently proposed post-hoc approach has been developed with neuroimaging models in mind, and the authors have shown [Friston et al., 2010a], using stochastic DCMs (see Chapter 1), that there is a very good agreement between the optimised and post-hoc model evidences.

This chapter explores the relation between optimised and post-hoc approximations to the model evidence for deterministic (rather than stochastic) DCM. Currently, deterministic DCM is the more standard methodology for making inferences about brain connectivity from functional magnetic resonance imaging (fMRI) data. We also investigate the effect of SNR on model comparison.

Since the main goal of DCM is to make inferences on the connectivity parameters we also compare the estimates of these parameters obtained with these two approaches. To this end we use synthetic and real fMRI data from a previously published study on attention to visual motion [Buchel and Friston, 1997]. This is the same dataset used in Friston et al. [2010a]. Although we use fMRI data, the methods described here can also be applied to other data modalities and statistical models, as long as the models that are compared are nested.

This chapter is structured as follows. In the Materials and Methods section (6.2) we review DCM for fMRI. We then focus on model optimisation and different approaches to estimate the model evidence and connectivity parameters. We then present (Section 6.3) and discuss (Section 6.4) results from comparing these approaches using synthetic and real fMRI data.

6.2 Materials and methods

In this section we briefly review DCM. We then look at different proxies for the model evidence. In particular we focus on a computationally efficient post-hoc approximation. In addition, we compare the estimates for the connectivity parameters obtained with this approach to the estimates obtained by optimising all models.

6.2.1 Dynamic causal modelling

As described in Chapter 1, DCM is a mathematical framework to estimate, and make inferences about, the coupling among brain areas and how this coupling is influenced by changes in experimental context [Friston and Penny, 2003]. It uses differential equations to describe the neuronal activity of interacting cortical regions and a forward model of how this neuronal activity is transformed into an observed response. This framework has been applied to fMRI, EEG and MEG [Kiebel et al., 2009], as well as LFP data [Moran et al., 2009]. Here, we focus on fMRI but the methods described below can also be applied to other data modalities.

Here we consider DCMs for fMRI that employ a deterministic bilinear model for the dynamics at the neuronal level (neurodynamics) and an extended Balloon model for the hemodynamic level. For non-linear, two-state or stochastic DCMs see Stephan et al. [2008], Marreiros et al. [2008a], and Friston et al. [2010a], respectively. The deterministic bilinear neurodynamics are described by the following multivariate differential equation:

$$\dot{z}(t) = (A + \sum_{j=1}^{M} u_j(t)B^j)z(t) + Cu(t), \tag{6.1}$$

where the dot notation denotes the time derivative. The variable z describes changes in neuronal activity resulting from the sum of three effects. First, the matrix

A encodes direct, or fixed, connectivity between pairs of regions. The elements of this connectivity matrix are not a function of the input and can represent both unidirectional and bidirectional connections. Second, the elements of B^j represent the changes in connectivity induced by the inputs u_j. M is the number of inputs. These condition-specific modulations or bilinear terms are usually the interesting parameters. Third, the matrix C encodes the direct influence of each exogenous input u_j on each area (Figure 1.8).

The overall structure of fixed, A, modulatory, B, and input, C, connectivity matrices constitutes our assumptions about model structure. This in turn represents a scientific hypothesis about the structure of the large-scale neuronal network mediating the underlying cognitive function.

As mentioned above, DCM for fMRI uses the extended Balloon model to describe how changes in neuronal activity give rise to the observed fMRI signals for each region. The full derivation of the model equations can be found in Buxton et al. [1998] and Friston et al. [2000]. In brief, see Eq. 5.3 (Chapter 5), for a particular region, neuronal activity, z, causes an increase in a vasodilatory signal, s, that is subject to auto-regulatory feedback. Inflow, f, responds in proportion to this signal with concomitant changes in blood volume, ν, and deoxyhemoglobin content, q, (Figure 1.6).

The hemodynamic parameters comprise the rate constant of the vasodilatory signal decay, τ_s, the rate constant for autoregulatory feedback by blood flow, τ_f, transit time, τ_0, Grubb's vessel stiffness exponent, α, and the resting oxygen extraction fraction, E_0. In this chapter, for identifiability reasons, only two of these parameters are estimated from the data for each region: $h = \{\tau_s, \tau_0\}$. The others are set to $\tau_f = \alpha = E_0 = 0.32$.

The blood oxygenation level dependent (BOLD) signal, y is then taken to be a static nonlinear function that comprises a volume-weighted sum of extra- and intravascular signals, Eq. 5.6 (see Chapter 5). See also Stephan et al. [2007a] for recent work on this BOLD signal model.

The parameters, θ, for a bilinear DCM, indexed by m, comprise the connectivity matrices as well as the hemodynamic parameters, i.e. $\theta = \{A, B, C, h\}$. The priors, $p(\theta|m)$, on both the connectivity and hemodynamic parameters are described in the Appendix. In current implementations of DCM, independent of modality, the model parameters are estimated from the data, y, using Bayesian methods, and models are compared using the model evidence.

6.2.2 Model evidence

The method of choice to approximate the model evidence for DCMs has been the variational free energy approximation [Friston et al., 2007a; Stephan and Friston, 2010], as described in Chapter 2. This method involves individually fitting (optimising) each model to data and then approximating the model evidence with a free energy bound. We refer to this approximation as the optimised model evidence. In contrast, Friston and Penny [2011] have proposed a post-hoc approximation to the evidence, which is computed by optimising only the largest of a set of models. This approach can be viewed as a generalisation of the well-known Savage-Dickey ratio [Dickey, 1971] (see Figure 6.1). In addition to the model evidence, the post-hoc scheme also provides estimates of the parameters for all reduced models from the full (optimised) model. Below we describe the two different approaches to approximate the model evidence and parameters (*optimised* and *post-hoc* approximations).

6.2.2.1 Optimised evidence

As mentioned in Chapter 2, VL updates the moments of an approximate posterior density, $q(\theta|y,m)$ by maximising an approximation to the negative variational Free Energy (henceforth 'free energy', F_m). This provides an approximation to the log model evidence, $\log p(y|m)$ [Penny, 2012].

Model comparison can then proceed using F_m as a surrogate for the log-evidence. We call this approximation *optimised evidence* because it comes out of the optimisation scheme described in Chapter 2, Section 2.2. This approximation to the model evidence is based on inverting all models in the model space. This is feasible only in a hypothesis driven procedure in which the model space comprises a small number of models. In large model spaces, optimising all models to obtain the evidences rapidly becomes computationally infeasible. For instance, in more exploratory analyses, one might be interested in looking at most, if not all, the possible connections and modulatory effects. The model space in this case can easily have thousands or millions of different networks. Below, we describe a less computationally expensive alternative to compute the model evidences.

6.2.2.2 Post-hoc evidence

This approach provides the model evidence and parameters for any nested (reduced) model within a larger (full) model as a function of the posterior density of the full

model [Friston and Penny, 2011]. This is a flexible approach that allows for post-hoc model selection without the need to invert more than a single model. In DCM the full model may be, for example, the fully connected network and the reduced models would correspond to networks with a sparser connectivity contained within this larger model.

The method assumes only the existence of a full model, $m_F \in M$, which shares the same likelihood with the set of reduced models, $m_i \in M$ and $\forall_i : m_i \prec m_F$:

$$p(y|\theta, m_i) \quad = \quad p(y|\theta, m_F), \tag{6.2}$$

This means the reduced models are constructed from the full model only by changing the priors on certain parameters $\theta^u \subset \theta$ as described below. This also implicitly assumes that the hyperparameters describing observation noise levels, λ_{obs}, are the same for the full and reduced models. This is not the case for the optimised model evidence approach, where λ_{obs} are optimised for each model.

We can then use Bayes rule to transform the above equality, Eq. 6.2. By re-arranging the terms we can write the ratio of model evidences in terms of the posterior and priors of the full and reduced model:

$$\frac{p(y|m_i)}{p(y|m_F)} \quad = \quad \frac{p(\theta|y, m_F)}{p(\theta|y, m_i)} \frac{p(\theta|m_i)}{p(\theta|m_F)} \tag{6.3}$$

Friston and Penny [2011] consider Eq. 6.3, under the Laplace approximation, as mentioned in Chapter 2. Under this approximation the posteriors, q, and priors, p, of the full and reduced models are Gaussian densities:

$$\begin{aligned} q(\theta|m_i) &= N(\mu_i, C_i) : C_i = P_i^{-1} \\ q(\theta|m_F) &= N(\mu_F, C_F) : C_F = P_F^{-1} \\ p(\theta|m_i) &= N(\eta_i, \Sigma_i) : \Sigma_i = \Pi_i^{-1} \\ p(\theta|m_F) &= N(\eta_F, \Sigma_F) : \Sigma_F = \Pi_F^{-1}, \end{aligned} \tag{6.4}$$

where $\eta_{i,F}$ and $\Pi_{i,F}$ are the prior means and precisions for the reduced (i) and full model (F), while $\mu_{i,F}$ and $P_{i,F}$ are the posterior means and precisions. Making use of

the assumptions of Eq. 6.4 in Eq. 6.3 the log model evidence for any reduced model can be written as a simple analytic function of the means and precisions of the prior and posterior of the full and reduced model:

$$\begin{aligned} F_i &= \log p(y|m_i) \\ &= \frac{1}{2}\log\frac{|\Pi_i||P_F|}{|P_i|||\Pi_F|} - \frac{1}{2}(\mu_F^T P_F \mu_F + \eta_i^T \Pi_i \eta_i - \eta_F^T \Pi_F \eta_F - \mu_i^T P_i \mu_i) + F^F. \end{aligned} \quad (6.5)$$

This is useful because the requisite means and precisions of the reduced model can be derived in a straightforward way from the means and precisions of the full model (see below).

The post-hoc approach can also be viewed as a generalisation of the Savage-Dickey density ratios [Dickey, 1971], in which the reduced models have certain parameters fixed at zero. To obtain these ratios we integrate Eq. 6.3 over the parameters. To do this we first partition the parameter space into two subsets of parameters $\theta = \{\theta^u, \theta^c\}$. The subset $\theta^u \subset \theta$ contains all the parameters which differ between the full, F, and reduced model, i. The remaining parameters θ^c are shared between the models, with equal priors: $p(\theta^c|m_i) = p(\theta^c|m_F)$. We refer to θ^u and θ^c as the unique and common parameters, respectively, and assume the priors factorise, i.e. $p(\theta|m_i) = p(\theta^u|m_i)p(\theta^c|m_i)$. With this notation, we can write Eq. 6.3 as follows:

$$\begin{aligned} \int p(\theta|y,m_i)\frac{p(y|m_i)}{p(y|m_F)}d\theta &= \int p(\theta|y,m_F)\frac{p(\theta|m_i)}{p(\theta|m_F)}d\theta \\ \frac{p(y|m_i)}{p(y|m_F)} &= \int\int p(\theta|y,m_F)\frac{p(\theta|m_i)}{p(\theta|m_F)}d\theta^u d\theta^c, \end{aligned} \quad (6.6)$$

where $\int p(\theta|y,m_i)d\theta = 1$. If we then use $p(\theta^u,\theta^c|y,m_F) = p(\theta^c|\theta^u,y,m_F)p(\theta^u|y,m_F)$ and the fact that the priors over θ^c are the same for both models we obtain the following result:

$$\begin{aligned} \frac{p(y|m_i)}{p(y|m_F)} &= \int\int p(\theta^c|\theta^u,y,m_F)p(\theta^u|y,m_F)\frac{p(\theta^u|m_i)}{p(\theta^u|m_F)}d\theta^u d\theta^c \\ &= \int p(\theta^u|y,m_F)\frac{p(\theta^u|m_i)}{p(\theta^u|m_F)}d\theta^u. \end{aligned} \quad (6.7)$$

When the reduced prior is a point mass (delta function), $p(\theta^u|m_i) = \delta(\bar{\theta}^u)$, that fixes the subset of parameters θ^u to a particular value, $\bar{\theta}^u$, the last equation, Eq. 6.7, reduces to the Savage-Dickey ratio (usually considered when $\bar{\theta}^u = 0$):

$$\frac{p(y|m_i)}{p(y|m_F)} = \frac{p(\theta^u = 0|y, m_F)}{p(\theta^u = 0|m_F)}. \tag{6.8}$$

This ratio has a simple intuitive interpretation: if we believe it is more likely that parameters θ^u are zero after seeing the data than before, then $p(y|m_i)/p(y|m_F) > 1$ and we have evidence in favour of the reduced model m_i. This is depicted in Figure 6.1.

The posterior of the full model can be obtained using the VL optimisation scheme, $q(\theta|y, m_F)$, described in Chapter 2. Again, under Gaussian assumptions we can write the previous ratio, Eq. 6.8, as follows:

$$\begin{aligned} F_i^u &= \log p(y|m_i) \\ &= \frac{1}{2}\log\frac{|P_F^u|}{|\Pi_F^u|} - \frac{1}{2}(\mu_F^{u^T} P_F^u \mu_F^u - \eta_F^{u^T} \Pi_F^u \eta_F^u) + F_F^u. \end{aligned} \tag{6.9}$$

This analytic formula is a special case of the post-hoc approach, Eq. 6.9, to calculate the model evidence of any reduced model as a function solely of the posterior mean and precision of the full model. The difference between Eq. 6.5 and Eq. 6.9 is the absence of quantities from the reduced model and the fact that all means and precisions are taken only for the subset of unique parameters, θ^u, which are not allowed to vary in the reduced model.

6.2.3 Post-hoc parameters

Once the full model has been optimised, Eq. 6.5 can be used to compute the model evidences for all reduced models from the full model. This results from the fact that, as we describe in the following, the posterior mean and precision of the reduced model parameters can also be determined from the mean and precision of the full model.

To obtain these estimates we again assume that the models differ only in the specification of the priors, i.e. they share the same likelihood, Eq. 6.2. Using this assumption we can subtract the linearised approximation to the posterior precision,

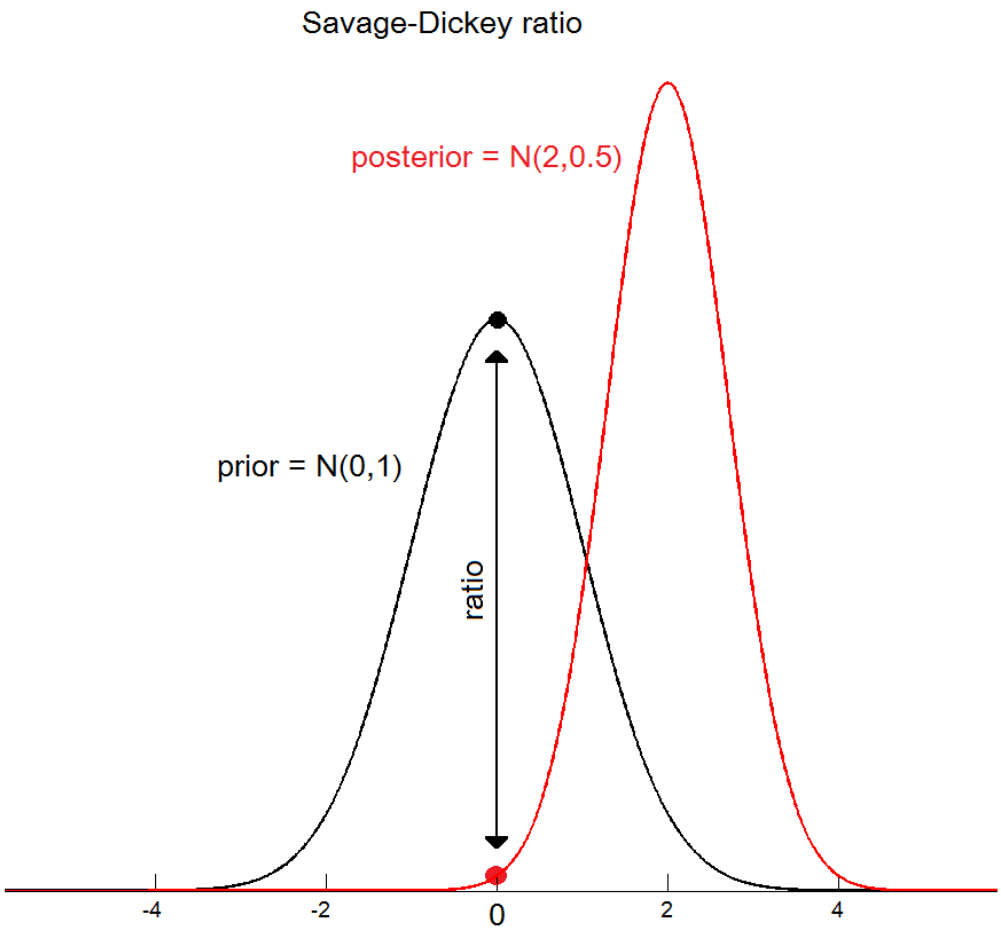

Figure 6.1: Savage-Dickey density ratio: This ratio is calculated by dividing the value of the posterior distribution over the parameters for the full model evaluated at $\theta = 0$, $p(\theta = 0|y, m)$, by the prior for the same model evaluated at the same point, $p(\theta = 0|m)$. These quantities are shown here for the case of a univariate Gaussian prior and posterior. The interpretation is simple: if it is less likely that parameters θ equal 0 after seeing the data (posterior) than before (prior), then $p(y|m_i)/p(y|m_F) < 1$ and we have evidence in favour of the full model, m_F, and vice-versa.

Eq. 2.5, of the full model from the precision of the reduced model and eliminate the terms that do not depend on the priors, such as $J^T C_\epsilon^{-1} J$. These terms are the same for all models and therefore cancel out in the subtraction. This yields the following result:

$$\begin{aligned} P_i - P_F &= J_i^T C_\epsilon^{-1} J_i + \Pi_i - J_f^T C_\epsilon^{-1} J_f - \Pi_F \\ &= \Pi_i - \Pi_F, \\ P_i &= P_F + \Pi_i - \Pi_F. \end{aligned} \tag{6.10}$$

Following exactly the same procedure we obtain the posterior mean of the reduced model as a function of the mean of the full model and the priors for both models. To summarise, the post-hoc approach provides estimates of the parameters (means and precision) under the Laplace assumption for any reduced model that can be obtained by inverting only the full model:

$$\begin{aligned} P_i &= P_F + \Pi_i - \Pi_F \\ \mu_i &= C_i(P_F \mu_F + \Pi_i \eta_i - \Pi_F \eta_F). \end{aligned} \tag{6.11}$$

This method is exact for linear models [Friston and Penny, 2011] (assuming that the likelihoods are the same, i.e. the models have a common noise variance). In the results section we test the validity of this approximation for bilinear deterministic DCMs. We compare the parameter estimates obtained with the post-hoc approach to the variational estimates obtained from optimising all models, using synthetic and real fMRI data.

Finally, once the model evidence and parameters have been estimated for each model, m, using the optimised or post-hoc approximations, these estimates can then be used for model selection, using Bayes Factors, Eq. 2.20, and posterior model probabilities, Eq. 2.10, as described in Chapter 2.

In the following section, we describe the dataset used to evaluate the methods presented here.

6.2.4 Subjects and task

The data were acquired by Buchel and Friston [1997] for one subject during an attention to visual motion paradigm. This dataset has been used to illustrate the post-hoc model selection approach on stochastic DCMs [Friston and Penny, 2011], as well as other methodologies from psychophysiological interactions [Friston et al., 1997] to Generalised Filtering [Friston et al., 2010b]. This dataset is publicly available on the SPM website (http://www.fil.ion.ucl.ac.uk/spm/). In this chapter we use 'DCM10' as implemented in SPM8, revision 4010.

Four consecutive 100 scan sessions were acquired, comprising a sequence of ten scan blocks of five conditions. The first was a dummy condition to allow for magnetic saturation effects. In the second, *Fixation*, the subject viewed a fixation point at the centre of a screen. In an *Attention* condition, the subject viewed 250 dots moving radially from the centre at 4.7 degrees per second and was asked to detect changes in radial velocity. In *No attention*, the subject was asked simply to view the moving dots. In a *Static* condition, the subject viewed stationary dots. The order of the conditions alternated between *Fixation* and visual stimulation (*Static*, *No Attention*, or *Attention*). In all conditions the subject fixated the centre of the screen. No overt response was required in any condition and there were no actual changes in the speed of the dots.

6.2.5 fMRI acquisiton and analysis

fMRI data were acquired with a 2 Tesla Magnetom VISION (Siemens, Erlangen) whole body MRI system, during a visual attention study. Contiguous multi-slice images were obtained with a gradient echo-planar sequence (TE=40 ms; TR=3.22s; matrix size=$64\times64\times32$, voxel size $3\times3\times3$mm). The data were pre-processed and analysed using the conventional SPM analysis pipeline (http://www.fil.ion.ucl.ac.uk/spm/), as described in Buchel and Friston [1997].

For this work we chose three representative brain regions defined as clusters of contiguous voxels in an 8 mm sphere surviving an F-test for all effects of interest at $p < 0.001$ (uncorrected), using SPM (see Chapter 1). These regions are: the primary visual cortex (V1), [0, -93 18] mm in MNI space, the middle temporal visual area (V5), [-36, -87, -3] mm, and the superior parietal cortex (SPC), [-27 -84 36] mm, [Buchel and Friston, 1997]. The activity of each region was summarised with its principal eigenvariate to ensure an optimum weighting of contributions from each

voxel within the region of interest (ROI).

6.3 Results

In this section we compare the optimised and post-hoc model evidences and parameter estimates with synthetic and the real fMRI dataset described above.

6.3.1 Synthetic data

Model space Model space comprised 128 models. These models have full fixed connectivity (bidirectional connections) between V1 and V5 and between V5 and SPC (Figure 6.2a). We allowed Motion to modulate only the connection from V1 to V5, but Attention was allowed to modulate any connection in the network, including the three self-connections (one for each region). In total we have 7 connections that can be modulated by Attention (3 self-connections + 4 fixed connections) resulting in $2^7 = 128$ different models. The full model (Figure 6.2a) is the model for which Attention modulates all these 7 connections.

We note that we chose to specify different models by changing only modulatory parameters because these connections comprise the bilinear terms (B matrices) in Eq. 6.1. This way we can evaluate Eq. 6.11, which provides estimates for the reduced parameters based on the full model, under non-linear conditions.

We started by generating data from model 96 by integrating the DCM equations (see Section 6.2.1). and adding Gaussian noise corresponding to a Signal to Noise Ratio (SNR) of 2.6 (data and noise had a standard deviation of about .350 and 0.135, respectively, SNR = 0.350/0.135 = 2.59), as used in Friston et al. [2010a]. In this model, Attention only modulates the connection between V1 and V5. Therefore, we refer to this model as the Forward model (Figure 6.2b). Figure 6.2c shows another example model, in which Attention modulates the connection between SPC and V5. We refer to this model as the Backward model.

To obtain the model evidence and parameter estimates for all 128 models using the optimised approach we had to invert (optimise) all these models. The optimisation procedure took approximately 5 hours in a 64-bit workstation. In comparison, for the post-hoc approach we only had to invert the full model, which took less than 2 minutes.

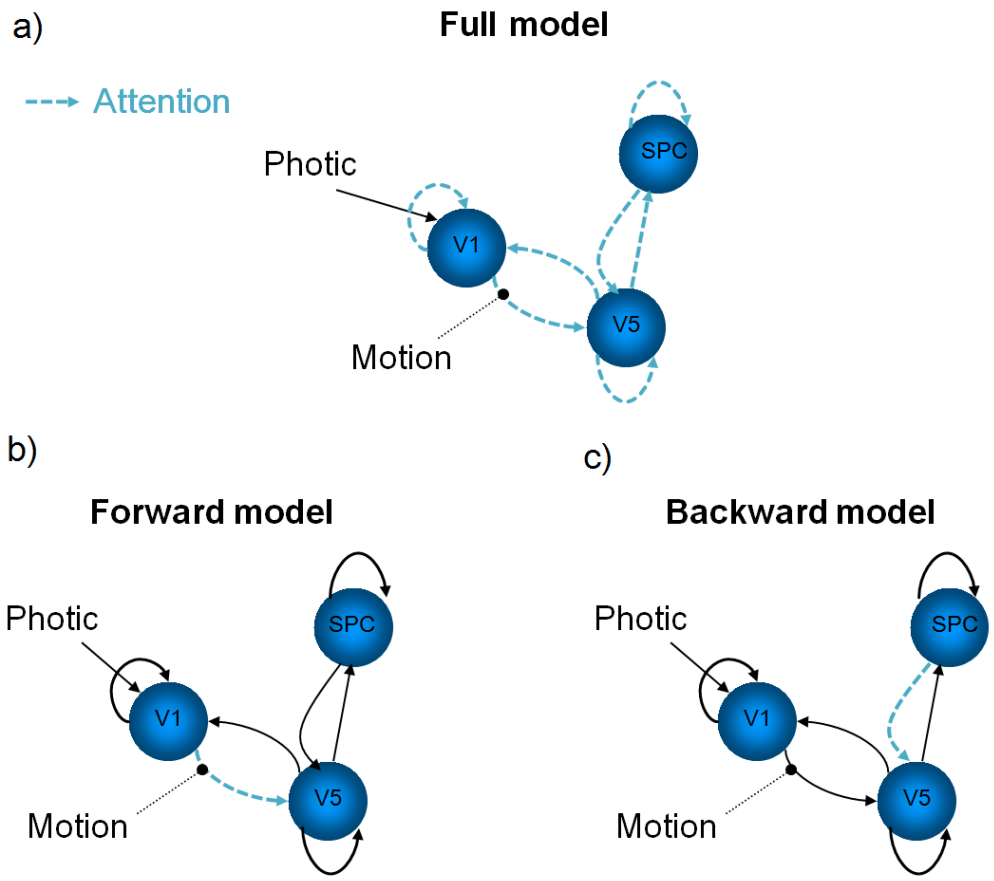

Figure 6.2: Models: a) Full model. In this model Attention modulates all the fixed connections and self-connections. This is the only model that needs to be inverted in order to estimate the evidence and parameters of all $2^7 = 128$ models, when using the post-hoc approximation. The following models vary in which connections are modulated by Attention (dashed arrows); b). In this model Attention only modulates the connection from V1 to V5. Consequently, we call it the Forward model (as opposed to the Backward model); c) Backward model: in this model Attention modulates the connection from SPC to V5.

Model evidence Figure 6.3a shows the optimised model evidence plotted against the post-hoc evidence for all 128 models. Here the evidence is relative to the worst model. As can be seen, the post-hoc measures correlate extremely well with the estimates obtained from optimising all models (they lie along the $y = x$ line). The actual correlation value is almost 1 ($r \approx 1$, p-value $< 1e^{-308}$). Figure 6.3b shows the relative evidences for the two approaches but as a function of graph size (number of edges). Again, the estimates for the model evidence obtained using the two approaches are extremely similar. Reassuringly, the true model (Forward model) has the highest log-evidence for both approximations and for the correct graph size (full circle): only one connection being modulated, in this case from area V1 to area V5.

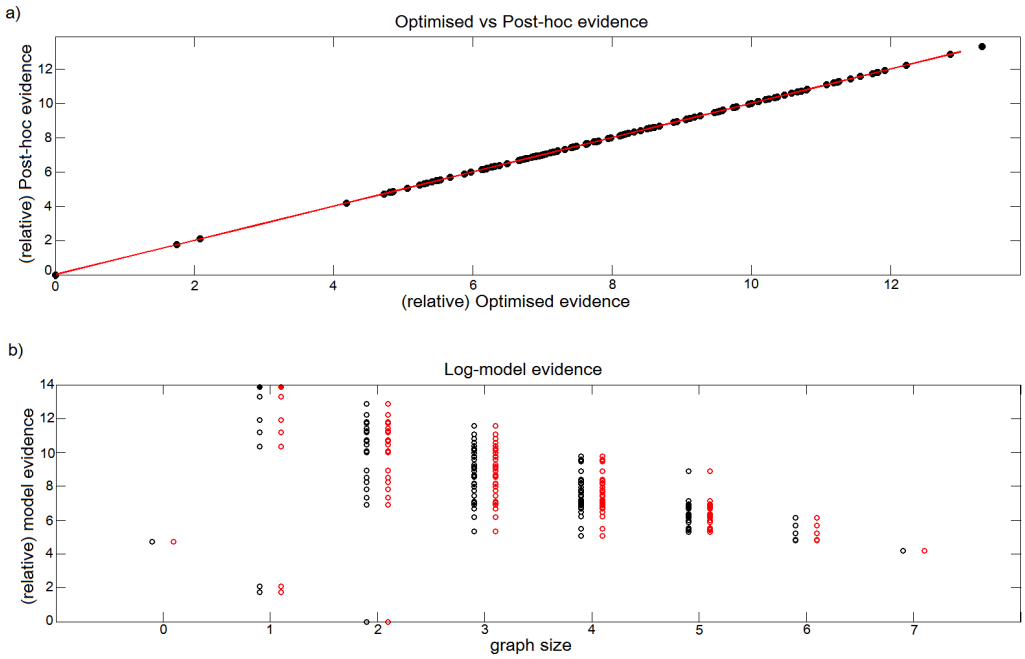

Figure 6.3: Synthetic data: a) Optimised log-model evidence (relative to worst model) versus post-hoc log-model evidence (128 synthetic models); b) Same data but plotted as a function of graph size (number of edges or modulated connections). The red circles correspond to the post-hoc estimates, while the black correspond to the optimised approach. The full circles indicate the best models for each approximation.

Using the same synthetic data generated from the Forward model (Figure 6.2b, also indexed as model 96) we looked at the model posterior probabilities for all 128 models. Again for the optimised approach we inverted all models, whilst for the post-hoc approach only the full model was inverted. As can be seen in Figure 6.4, even

though, as expected due to the number of models [Penny et al., 2010], the posterior mass is diluted over the models and no single model has very high probability, the true model (marked by the asterisk) has the highest posterior in both the optimised and post-hoc approaches.

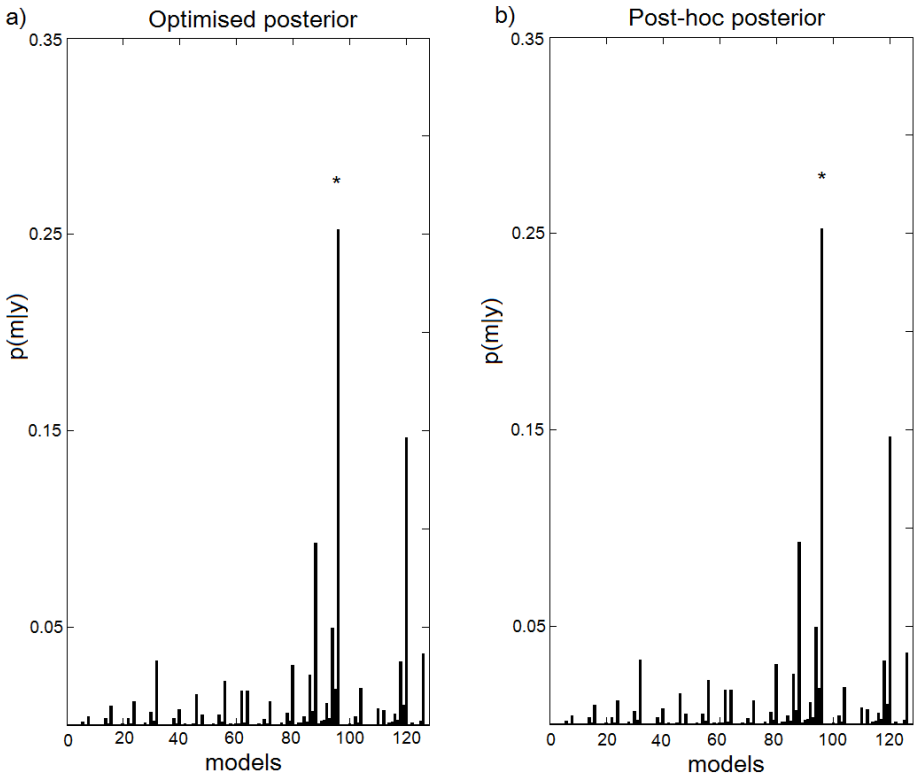

Figure 6.4: Synthetic data - Bayesian model selection: a) Optimised model posteriors. The data were generated from model 96, Forward model (Figure 6.2b), (marked by an asterisk, $*$). This model is also the best model for both approximations; b) Post-hoc posterior probabilities. The backward model is model number 126 and the Bayes Factor between the Forward and Backward model is 1.94 (as expected from Figure 6.6).

Model parameters We then looked at the connectivity parameter estimates obtained with the optimised and post-hoc estimation approaches. Figure 6.5a shows the true connection strengths that were used to generate the data, again from the same model (Forward model). We have 7 connections but only one of them (from V1 to V5) has a value different from zero. The second row of plots in Figure 6.5 shows the parameter estimates (mean and 95% confidence intervals) obtained with

the optimised and post-hoc approaches, respectively, corresponding to the best model identified previously (Figure 6.4). As can be seen, both approaches identify the second parameter as being the only connection significantly different from zero. The true parameter value is 0.23 and both the optimised and post-hoc posterior means for this parameter are estimated as 0.29. The parameter estimates are summarised in Table 6.1.

Table 6.1: Parameter estimates: posterior mean and 95% confidence intervals of the best model obtained with the optimised and post-hoc methods for synthetic and real data (first and second row of results for each connection, respectively). The subscript *op* means optimised, and *ph* means post-hoc.

Parameter estimates				
Data	Connection	μ_{true}	μ_{op}	μ_{ph}
Synthetic	V_1	0	0.00 ± 0.00	0.00 ± 0.00
Real		-	0.80 ± 0.45	0.80 ± 0.45
	$V_1 \rightarrow V_5$	0.23	0.29 ± 0.10	0.29 ± 0.10
		-	1.14 ± 1.08	1.14 ± 1.08
	$V_5 \rightarrow V_1$	0	0.00 ± 0.00	0.00 ± 0.00
		-	-0.79 ± 0.52	-0.79 ± 0.52
	V_5	0	0.00 ± 0.00	0.00 ± 0.00
		-	0.85 ± 0.96	0.85 ± 0.96
	$V_5 \rightarrow SPC$	0	0.00 ± 0.00	0.00 ± 0.00
		-	0.00 ± 0.00	0.00 ± 0.00
	$SPC \rightarrow V_5$	0	0.00 ± 0.00	0.00 ± 0.00
		-	-2.79 ± 1.16	-2.79 ± 1.16
	SPC	0	0.00 ± 0.00	0.00 ± 0.00
		-	0.00 ± 0.00	0.00 ± 0.00

These results show that, even though Eq. 6.11 is only an approximation in the case of non-linear models, it provides excellent estimates for bilinear DCMs.

Signal-to-noise ratio The previous results have been obtained by generating data from one model and looking at how the different approaches to estimate the evidence and parameters compare using a fixed SNR similar to the SNR of the real fMRI dataset. This dataset comes from a block design paradigm and therefore has relatively high SNR (see Chapter 1). In this section we explore the behaviour of the two approaches for different values of SNR. To this end we performed two different model comparisons: i) we generated data from the Forward model and compared this to

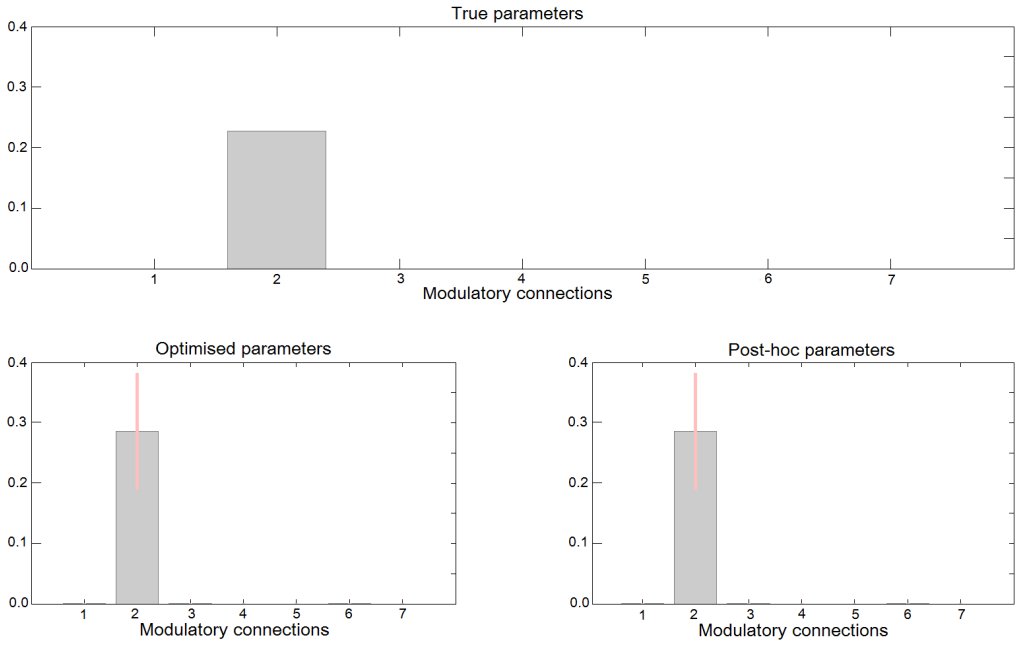

Figure 6.5: Synthetic data - parameter estimates: a) True parameters from which the data were generated; Only the second parameter is modulated: forward connection from V1 to V5; b) Optimised and post-hoc parameter estimates for the best model (Figure 6.4). The error bars correspond to 95% confidence intervals. The parameters 1 to 7 (x axis) correspond to the 7 connections possibly modulated by Attention.

the Backward model (Figure 6.2); ii) we generated data from the full model and compared this to the Forward model described above. For both these comparisons we varied the SNR of the data from 0.35 to 3.35 in intervals of 0.1.

We also repeated the data generation, optimisation and model comparison 10 times for each SNR, in order to have 10 realisations of the same result. We then plotted (Figure 6.6) the mean log-Bayes Factor and 95% confidence intervals for each comparison as a function of the SNR. To obtain these results with the optimised approach we had to invert both the Forward and Backward models (first comparison), and Full and Forward models (second comparison) for each SNR and realisation. For the post-hoc approach we had only to invert the Full model for each SNR and repetition in both cases.

Figure 6.6a shows that, as expected, the log-Bayes factors increase with higher SNR. However, our simulations suggest that the optimised approach seems to reach significant results (log-Bayes factor higher than 3) slightly faster than the post-hoc approach. This disparity between the two approaches might be due to the fact that in the optimised approach the hyperparameters, λ, are estimated for each model, whilst in the post-hoc approach these are assumed to be the same for all models (equal to the estimates for the full model).

The fact that the log-Bayes factors are positive (with increasing SNR) means that both methods are selecting the true model as the best model, with increasing confidence. One other thing to note is that the error bars are relatively smaller for the post-hoc approach, suggesting that the results for the optimised evidence are more inhomogeneous. This may be because the optimised approach reaches different local minima on different optimisation runs.

At low SNR (below 1) the log-Bayes factors are close to zero with the error bars enclosing this number, as expected. In this case none of the methods select a winning model. However, for very low SNR (first two points) both methods seem to slightly prefer the backward model (BF < 1) . This result might be due to the difficulty of estimating the models under very low SNR conditions, which can lead to inaccurate model selection results with both methods.

The results for the second comparison, where the true model is the full model, (Figure 6.6b) are very similar. The log-Bayes factors for the optimised approach increase significantly faster than the post-hoc approach, but the error bars are again slightly bigger. Here too the log-Bayes factors increase positively, which means that both methods are selecting the full model as the best model, even though this model is

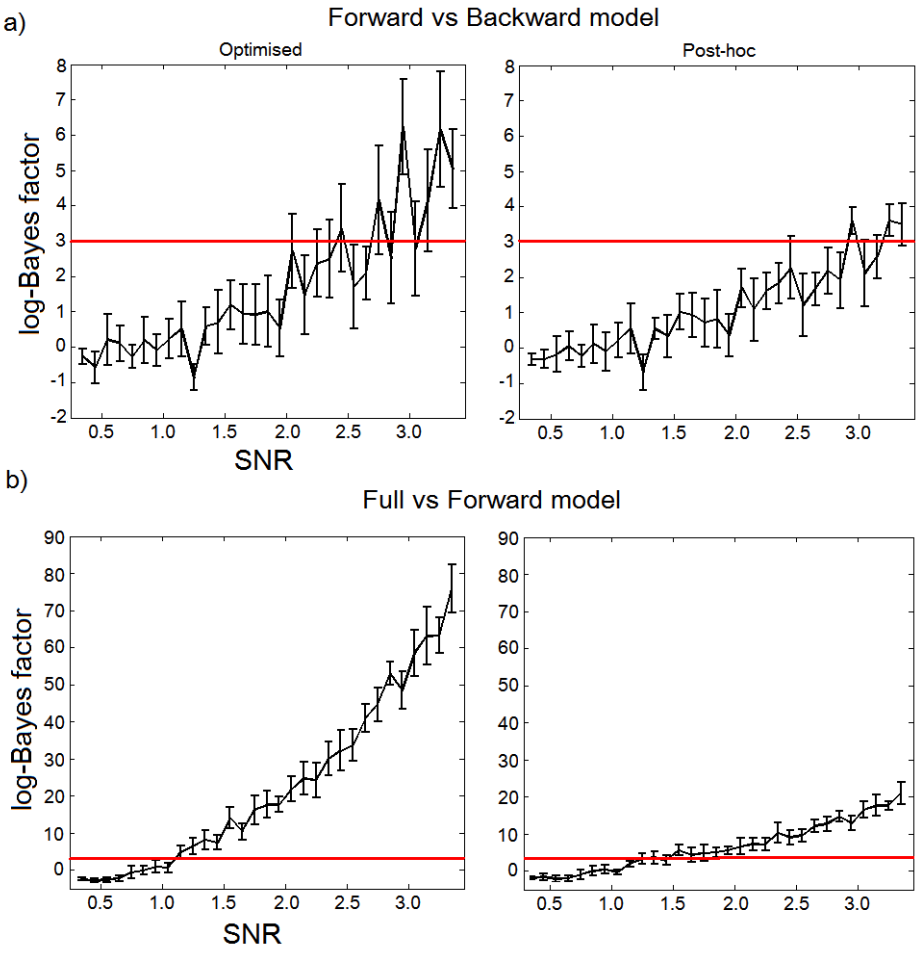

Figure 6.6: Signal-to-noise ratio: a) log-Bayes factors (between the forward and backward model) averaged over 10 repetitions of the same comparison (with 95% confidence intervals) as a function of the signal to noise ratio used to generate the data (from forward model); b) log-Bayes factors (between the full and forward model) averaged over 10 repetitions of the same comparison (with 95% confidence intervals) as a function of the signal to noise ratio used to generate the data (from full model). The red line corresponds to a log-BF of 3.

penalised for extra complexity. However, in the low SNR case (first 4 points, between 0.35 and 0.65) both methods seem to select the Forward model as the best model (negative Bayes factors). This means that in the almost complete absence of data (i.e. presence of high levels of noise), the full model is highly penalised and both model selection methods prefer the simpler hypothesis, the Forward model.

We then regressed the post-hoc evidences onto the optimised evidences and looked at the regression coefficients. In Figure 6.7 we plot these coefficients for both comparisons (Figure 6.7a and b)). As can be seen in the first case (Figure 6.7a) the regression coefficients are all significantly different from zero and seem to slightly increase as a function of SNR. In the full versus forward model case (Figure 6.7b) the results are very similar. Again all coefficients are significantly different from zero and increase as a function of SNR.

The previous results show that there is a linear relationship between the optimised and post-hoc measures (even in low SNR conditions) and that this relationship increases with increasing SNR.

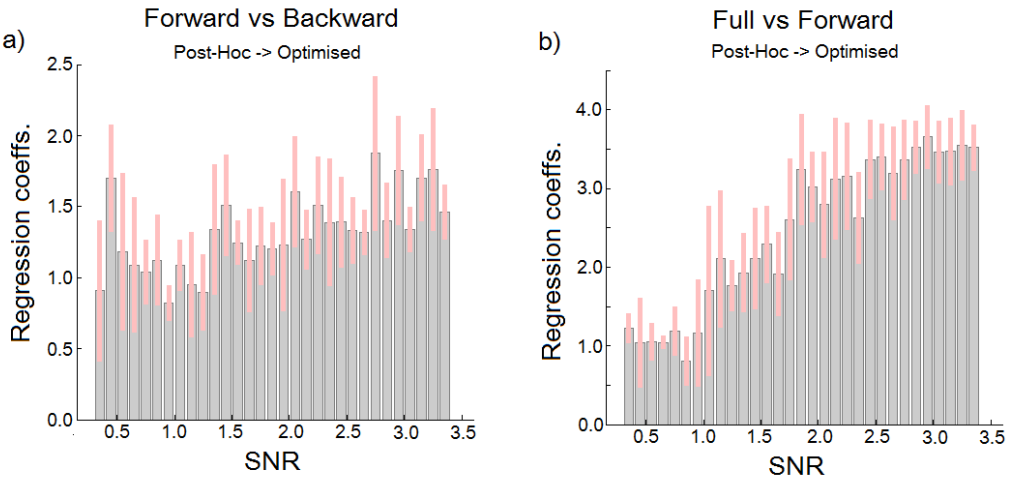

Figure 6.7: Signal-to-noise ratio: a) regression coefficients (and 95% confidence intervals) between optimised and post-hoc Bayes factors (comparing the forward model, true model, to the backward model) as a function of the signal to noise ratio; b) regression coefficients (and 95% confidence intervals) between optimised and post-hoc Bayes factors (comparing the full model, true model, to the forward model) as a function of the signal to noise ratio.

In summary, the results obtained with synthetic data show that both approximations to the model evidence presented here yield similar results but the post-hoc

approach reduced the computation time from a couple of minutes per model to a couple of seconds. In addition, even though the SNR of this dataset is relatively high (it is a block rather than event-related design) the post-hoc approach was also able to obtain the true model in lower SNR scenarios. The post-hoc estimates of the connectivity strengths were also very similar to the optimised and true estimates.

6.3.2 fMRI data

After testing the methods on synthetic data we turned to the fMRI dataset acquired by Buchel and Friston [1997]. Here we used the time-series from the three brain regions V1, V5 and SPC for one subject as described above.

Model space We used the same set of 128 models as defined before. The full model is the same full model used with synthetic data, in which Attention modulates all fixed connections between the three areas, as well as their three self-connections (Figure 6.2a). In the optimised approach all 128 models were fitted to the fMRI signals. This took roughly the same amount of time to fit the synthetic data, since we used a similar signal to noise ratio to the real data. In the post-hoc approach only the full model was fitted to the fMRI data. Again this approach computed the evidences for all models in a few seconds.

Model evidence We plotted the post-hoc evidences against the model evidence obtained with the optimisation approach. As suggested by the results obtained with synthetic data, these measures correlate extremely well with the optimised evidences for this dataset (Figure 6.8a), where $r \approx 1$ (p-value $< 1e^{-308}$). The best model identified by the optimised evidence is the same model (model 6) for the post-hoc approach. This model corresponds to a graph-size of 5, meaning that Attention modulates five connections (Figure 6.8b): self-connections of V1 and V5, plus connections from V1 to V5, V5 to V1, and SPC to V5.

Figure 6.9 shows the model posteriors obtained with both approaches for all 128 models using real fMRI data. As shown above (Figure 6.8b), both methods identify model 6 as the best model with posterior probability close to 0.16.

Model parameters The parameter estimates (means and 95% confidence intervals) for the best model (model 6) are very similar for both approaches (Figure 6.10).

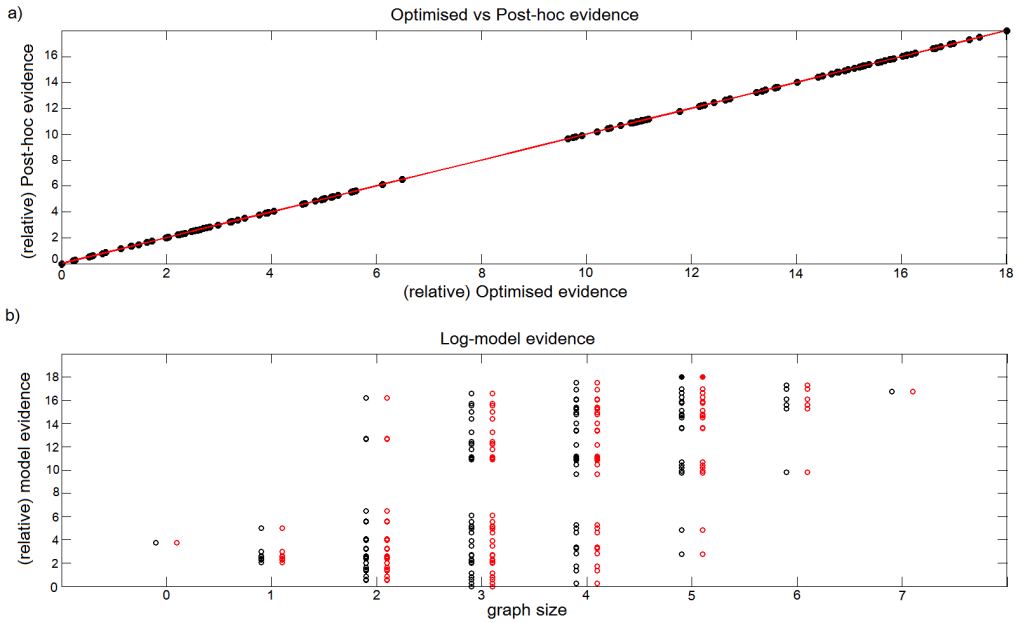

Figure 6.8: fMRI data: a) Optimised log-model evidence (relative to worst model) versus post-hoc log-model evidence (128 models); b) Same data but plotted as a function of graph size (number of edges or modulated connections). The red circles correspond to the post-hoc estimates, while the black correspond to the optimised approach. The full circles indicate the best models for each approximation.

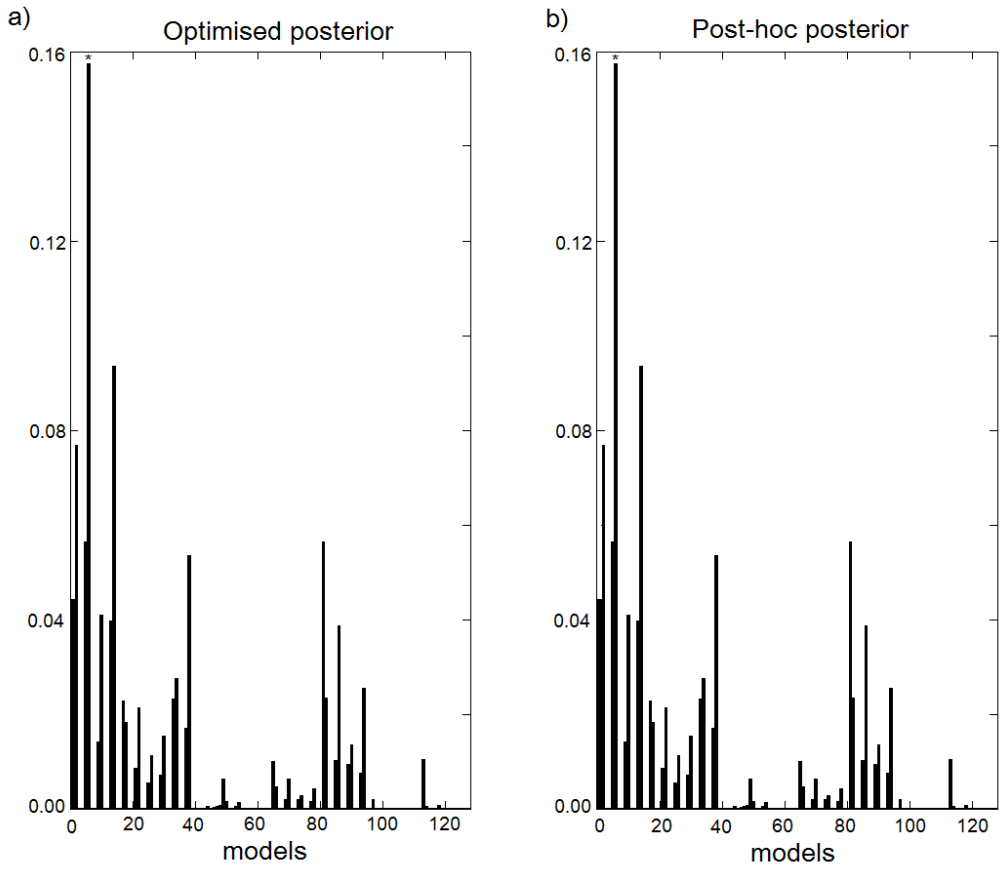

Figure 6.9: fMRI data - Bayesian model selection: a) Optimised model posteriors. The best model, model 6, is marked by an asterisk, $*$. b) Post-hoc model posteriors.

We can see that 5 of the total of 7 parameters seem to have values different than zero (although the error bars cross the zero line for the fourth parameter), as suggested by the best model by graph size in Figure 6.8b (graph size 5). The values estimated for each connection are summarised in Table 6.1.

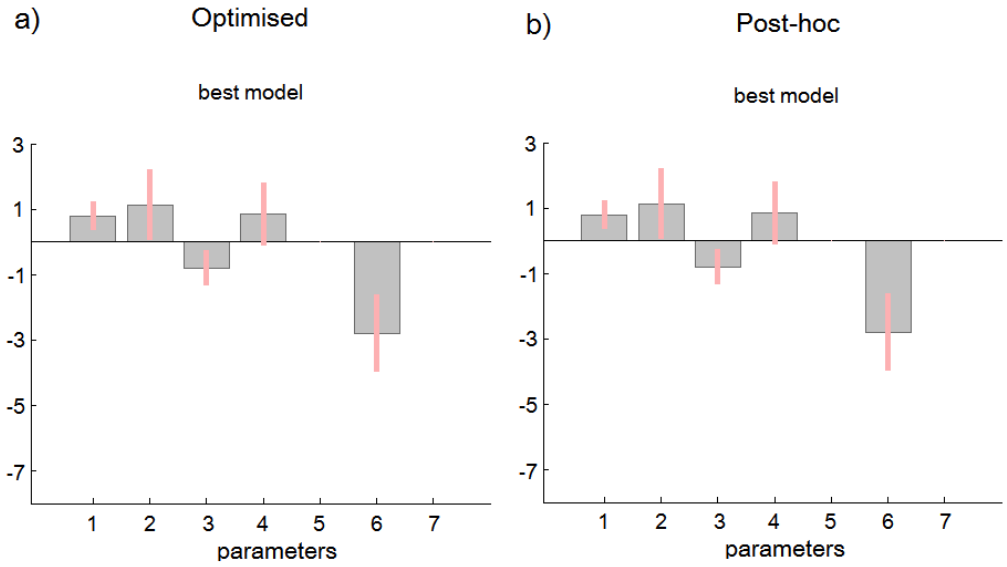

Figure 6.10: fMRI data - parameter estimates: Optimised and post-hoc parameter estimates for the best model, model 6, (Figure 6.9). The error bars correspond to 95% confidence intervals. The parameters 1 to 7 (x axis) correspond to the 7 connections possibly modulated by Attention.

In summary, the results obtained with the real fMRI dataset are very similar to the ones obtained for synthetic data. Again the optimised and post-hoc methods provide very similar results both for the evidences and model parameters.

6.4 Discussion

This chapter presents and evaluates a recent approach, post-hoc selection [Friston and Penny, 2011], for estimating the model evidence and parameters of deterministic DCMs. This method offers substantial computational advantages to the 'optimised' approach that is currently used [Friston et al., 2007a].

Our results show very good agreement between optimised and post-hoc model evidences, for both synthetic and real fMRI data. This suggests that the post-hoc

method offers valid estimates of the evidence with little computational cost. The post-hoc approach reduced the computation time needed to optimise and compare hundreds of models from several hours to a few minutes.

The post-hoc method also provides estimates of the model parameters. Here we found that the post-hoc and optimised approaches yield very similar results. We have also shown that the post-hoc approximation to the model parameters, which is exact for linear models, seems to be a good approximation for non-linear models, such as DCMs.

As an aside, we note that we have also compared the post-hoc approximation to the model evidence, Eq. 6.5, to the Savage-Dickey approximation, Eq. 6.9, which is a special case of the former. As expected, these two measures yielded numerically identical results, including identical model posteriors. Moreover, when we regressed the Savage-Dickey Bayes-factors onto the post-hoc Bayes-factors for a wide range of SNRs (same as described in Results), we obtained regression coefficients equal to 1 for all SNRs.

Although this post-hoc approach is very computationally efficient, the number of possible models to compare can rapidly explode when considering networks with many regions and all possible connections between them. In this case, it might be impossible to compute the evidences and parameters for all models and one might have to sample the space of models. For instance, Pyka et al. [2011] use genetic algorithms to accelerate model selection of large numbers of DCMs. Therefore, one extension to this work would be to use greedy searches and stochastic search algorithms that efficiently compute the post-hoc evidences and parameter estimates in arbitrarily large model spaces.

To conclude, our results provide evidence supporting the use of the post-hoc method proposed by Friston and Penny [2011] for model selection (and parameter inference) of bilinear deterministic DCMs.

Chapter 7

Discussion and conclusions

The previous four chapters described the novel methodologies and empirical work pursued in this thesis. This chapter presents a summary of this work and discusses its main implications for the field of imaging neuroscience. In addition, this chapter presents future research directions that can be followed from the work presented here, as well as a summary of original contributions and publications can that came out of this thesis.

7.1 Summary of work

The aim of this thesis was to develop and validate new model comparison techniques that can be used by the neuroimaging community to investigate brain function. In addition, this thesis applied existing comparison methods to study the relationship between neuronal activity and blood flow. This topic is known as neurovascular coupling and is still under intense debate.

This thesis addressed the neurovascular coupling problem by first looking at how changes in blood flow (measured with BOLD-fMRI) relate to the temporal dynamics of neuronal activity (measured with EEG) in the human brain (see Chapter 3). This work followed previous studies using combined EEG-fMRI data, where changes in the fMRI signals were regressed onto EEG oscillatory power. This chapter uses the same methodology but posits different models, or transfer functions, of how the spectral profile of neuronal activity relates to BOLD. These functions correspond to different features of the EEG power spectrum, such as the time-varying total power, time-varying power at a single or combination of frequency bands, and the relative

power between high and low frequencies. These functions were then compared using a model selection approach based on F-tests. Results suggest that changes in BOLD are indeed associated with changes in the spectral profile of neuronal activity and that these changes do not arise from one specific frequency band. Instead they result from the dynamics of the various frequency components together, in particular, from the relative power between high and low frequencies.

The model selection approach used in Chapter 3 is useful and provides valid comparisons, but is limited to nested models. This limitation lead us to develop a new Bayesian mapping approach for model selection at the group level (Chapter 4). This technique is based on the combination of an approach for providing log-evidence maps for each model and subject, with a random-effects approach for model selection. This procedure constructs PPMs for BMS inference. We illustrated the method using fMRI data from subjects performing a cued detection task. Results showed that this is a valid approach and allows neuroimagers to make inferences about regionally specific effects in the brain without being restricted to comparing nested hypotheses. It is therefore more general than F-tests and we have shown that it is also robust to outliers.

This thesis then focused again on neurovascular coupling, this time from a mechanistic point-of view. This perspective is not afforded by the approach used in Chapter 3, which was based on linear models (SPM framework) and classical F-tests for model comparison. Therefore, in Chapter 5, we used a detailed biophysical framework and posit models embodying biologically plausible hypotheses about neurovascular coupling. We adjudicated between hypotheses using Bayesian model evidence, and EEG-fMRI data from human visual cortex. We compared three models: the first assumes that blood flow depends on the amount of vasodilatory nitric oxide released as a result of input synaptic activity; the second assumes blood flow is driven by the firing rate (output spiking activity) of local pyramidal cells; these hypotheses are then compared against a third model where both these quantities contribute to the BOLD response. Results showed that the BOLD signal is dependent upon both synaptic and spiking activity but that the relative contribution of these two factors are dependent upon the underlying neuronal firing rate. When the underlying neuronal firing is low then BOLD signals are best explained by synaptic input, when the neuronal firing rate is high then both synaptic and spiking activity are required to explain the BOLD signal.

The models used in Chapter 5 are a special case of brain connectivity models, when

there is only one area. Connectivity models, such as DCMs, allow neuroimagers to infer how interactions between different brain regions are mediated. In recent years a trend has arisen for comparing models with many brain regions and all possible connections between them. With current methods it can rapidly become computationally infeasible to estimate all models for comparison. This fact has lead to the recent development of a post-hoc Bayesian approach that can be used to quickly score large numbers of connectivity (or any other nested) models. In the last chapter of this thesis (Chapter 6) we have applied this approach to deterministic DCMs for fMRI, which have become a standard method for effective connectivity analyses. We use data from an attention to visual motion study and showed that the post-hoc approach provides valid model evidence and parameter estimates, without having to invert more than a single (full) model. The advantage is therefore a huge reduction in computational time necessary for model selection.

7.2 Implications of work

This section discusses the main implications of this work for the field of imaging neuroscience.

Chapter 3 This chapter has shown that it is possible to disambiguate between models of how BOLD relates to underlying neuronal activity in the human brain, using the SPM framework and simultaneous EEG and fMRI data. Results from this chapter have therefore helped clarifying our current understanding of how BOLD relates to the frequency dynamics of neuronal activity, as measured with EEG. These results are important not only for our comprehension of neurovascular coupling but also for improving the interpretabiliy of fMRI findings. Knowing how fMRI relates to underlying temporal dynamics allows us to be more certain of what type of neuronal activity is measured with this technique and how this measure relate to observations obtained with EEG/MEG. In addition, the results in this chapter can be used to inform better forward models for both EEG and fMRI data, which can then be used to further investigate the neurovascular coupling, as was done in Chapter 5.

Chapter 4 The PPMs for BMS developed in this chapter allow one to compare any non-nested GLMs and therefore can be particularly useful for model-based fMRI [O'Doherty et al., 2007]. In model-based fMRI, signals derived from computational

models of specific cognitive processes are correlated against fMRI data from subjects performing a relevant task to determine brain regions where activity is consistent with that model. Differences between models are usually subtle and it can therefore be difficult to compare different hypotheses using F-tests. In this case, models need to be nested and one model might explain away most of the variance of the other models. The PPM approach developed in this chapter does not suffer from this limitation and can therefore facilitate model-based inferences of how particular cognitive processes are implemented in specific brain areas.

Chapter 5 Similarly to Chapter 3, this chapter has shown that it is possible to disambiguate between biophysical mechanisms for neurovascular coupling, using non-invasive techniques in the healthy human brain. The results obtained here are again important not only to understand what type of neuronal activity is measured by fMRI but also for relating fMRI findings to other measures, in particular, to electrophysiological data from invasive animal studies. The fact that, as shown in this chapter, BOLD seems to reflect mostly synaptic or both synaptic and spiking activity depending on the frequency regime, is also important for the interpretation of fMRI results under different stimuli conditions. For example, single cell electrophysiology measures firing rates, and therefore our results show that these measures are more similar to fMRI measures at high firing rates. In addition, the framework presented here has been developed so that it can be used efficiently with any EEG or MEG and fMRI dataset, where the temporal resolutions of the techniques are very different. This facilitates further research about neurovascular coupling using these types of models.

Chapter 6 has proved the validity and usefulness of a recently developed post-hoc approach for quickly scoring large number of deterministic DCMs. This work therefore opens the door for further exploratory analyses of brain networks, using DCMs and fMRI data, without the need for excessive computational power.

7.3 Directions for future research

This section puts forward ideas for potential future research based on the work developed in this thesis, and proposes suggestions for methodological advances.

Chapter 3 This chapter compared different coupling (transfer) functions, instantiated as linear models, from neuronal activity to BOLD, using simultaneous EEG-fMRI data. A concern with the model comparison approach taken in this chapter is that it is based on GLMs and F-tests, which restrict one to making inferences about nested models. If no natural nesting exists, then the regressors from all models are placed in the same design matrix and F-tests used to infer whether sets of variables explain additional variance. Whilst this approach is commonplace [Friston et al., 2007b], it is nonetheless suboptimal as compared to direct comparison of models using the Bayesian model evidence criterion, as discussed in Chapter 2. This has been shown by Penny et al. [2007] in the context of fMRI signals. In Chapter 4, we have extended Bayesian model comparison to imaging data from group studies. This method could therefore be used with EEG-fMRI data for comparing neurovascular coupling models in the human brain, such as the transfer functions compared in this chapter, without the restriction of the models having to be nested.

Another concern in the analyses we have presented here is in the use of EEG regressors as a surrogate for neuronal activity. This approach has previously been used by a number of groups [Goldman et al., 2002; Laufs et al., 2003; Lemieux et al., 2001; Moosmann et al., 2003]. In this chapter we followed the same rationale but additionally employed a visual flicker stimulation paradigm to elicit evoked activity in sensory cortex. We then used the first principal component of the EEG data to isolate activity that was primarily related to the stimulus paradigm. We note that this approach could be improved in a number of ways. First, one could employ multiple PCA or ICA components [Eichele et al., 2005, 2009; Vigario et al., 2000], which might better isolate activity from specific processes or brain regions. Second, one could use regressors derived from EEG source reconstructions as in Wan et al. [2006]. A problem with these approaches, however, is that they are no longer compatible with a whole-brain SPM analysis approach, as that requires the same design matrix at all voxels. They are nevertheless worth pursuing.

Another improvement to this work would be to use fMRI data recorded concurrently with intracranial EEG (iEEG). Although iEEG measurements in humans can only be obtained under very restricted circumstances, we expect it to play a major role in investigating neurovascular coupling, as this will provide more direct access to the various cortical and subcortical regions that have little impact on the scalp EEG. This may help, for example, to resolve to what extent, if at all, BOLD and EEG are differentially sensitive to endogenous lower frequency global states versus

higher frequency local processing [Laufs, 2008].

Chapter 4 This chapter presented the construction of PPMs allowing for BMS at the group level. These maps are produced by combining a model evidence mapping approach with an RFX approach for model selection. In the current implementation of this technique, log-evidence maps were smoothed by a user specified Gaussian kernel. The arbitrariness of having to specify a particular amount of smoothing could be eliminated by including a spatial model over the model frequencies. This way the amount of smoothing would be estimated from the data within the same Bayesian framework. This procedure mirrors corresponding developments in the analysis of group data from M/EEG source reconstructions [Litvak and Friston, 2008].

Another potential improvement in the methodology presented in this chapter would be to use the Savage-Dickey ratios or post-hoc selection approach, described in Chapter 6, to decrease the computational time necessary to fit the models and obtain the log-evidence maps. This would allow one to explore large model spaces (e.g. in the order of hundreds of models or more) using BMS maps.

Furthermore, we envisage that the methodology developed in this chapter will be used not only for model-based fMRI, but also for model-based EEG/MEG, where, similarly to fMRI, computational models can be fitted to behaviour measures, which can then be used as predictors of scalp or source-reconstructed data.

Chapter 5 This chapter used EEG-fMRI data and a biophysically informed mathematical model to investigate the relationship between neuronal activity and the BOLD signal, in human visual cortex. A natural extension to the work presented in this chapter is the inclusion of multiple cortical units in the model representing multiple brain areas. For instance, sub-cortical areas such as the thalamus and other cortical areas activated by the experimental task could be included. Having more than one area would facilitate the differentiation between input and local processing synaptic activity, such as in Sotero and Trujillo-Barreto [2008]. In a recent study, Harris et al. [2010] have decomposed the effect of these two types of synaptic activity on hemodynamic signals by reducing the thalamic input to a rodents cortex. The authors found that although both input and local neuronal processing contribute to BOLD signals, as previously found, this contribution is larger from local processing. This hypothesis could be tested within the framework proposed in this chapter.

Another extension to the work presented here would be to probe the contribution

of excitatory and inhibitory neuronal populations to the generation of BOLD signals, such as in Sotero et al. [2009]. This model-driven approach could, for instance, be used to study the findings of Boorman et al. [2010], where a negative BOLD response in deeper cortical layers, adjacent to positive-BOLD areas, was found to be associated with a reduction in local neuronal firing. Very recently, Lee et al. [2010] have optically driven genetically modified inhibitory cells and measured a negative BOLD signal in response to this stimulation, in the rat cortex. This result could inform the development of new generative models of neurovascular coupling.

Finally, in the long term, we anticipate that the modelling framework presented in this chapter will be used to test neurovascular coupling hypotheses in a variety of experimental contexts with a range of subject cohorts.

Chapter 6 This chapter validates a recently proposed Bayesian approach for quickly comparing large numbers of connectivity models, based on inverting a single model, using fMRI data and deterministic DCMs. Although this post-hoc approach is very computationally efficient, the number of possible models to compare can rapidly explode when considering networks with many regions and all possible connections between them. In this case, it might be impossible to compute the evidences and parameters for all models and one might have to sample the space of models. For instance, Pyka et al. [2011] use genetic algorithms to accelerate model selection of large numbers of DCMs. Therefore, one extension to this work would be to use greedy searches and stochastic search algorithms that efficiently compute the post-hoc evidences and parameter estimates in arbitrarily large model spaces.

Another extension of this work would be the application of the post-hoc approach to other modalities, such as DCMs for EEG and MEG data [Kiebel et al., 2009], as well as to other types of DCMs, such as non-linear [Stephan et al., 2008] and two-state DCMs [Marreiros et al., 2008b].

Finally, another possible application of this work would be to use the post-hoc approximation to the model evidence, described here, in a voxel-wise manner, to be able to create computationally efficient log-model evidence maps. These maps could then be used in conjunction with the methodology described in Chapter 4 to construct PPMs for BMS at the group level. The advantage would be an enormous reduction in computational time necessary to calculate the evidence maps.

7.4 Conclusions

The work presented in this thesis developed and applied different model selection techniques that can be used to investigate brain function, using imaging data. These techniques ranged from classical methods, based on F-tests (Chapter 3), to more sophisticated Bayesian approaches (Chapters 4 to 6).

One of the open research topics that this thesis focused on is the relationship between neuronal activity and BOLD, known as neurovascular coupling. The work presented here shows that it is possible to non-invasively investigate different hypotheses (models) for the coupling, both at a phenomenological (Chapter 3) and neurophysiological (Chapter 5) level, in human brain using combined EEG and fMRI data.

This thesis has also shown that it is possible to construct PPMs for BMS at the group level, which allow one to characterise regional specific effects using brain imaging data, without being restricted to comparing nested models. This approach was shown to be able to distinguish between models of how the brain integrates information over time, using fMRI data from a group of subjects (Chapter 4).

Finally, this thesis has shown that it is possible to compare large numbers of deterministic dynamic causal models by inverting a single model, using a post-hoc approximation to the model evidence (Chapter 6).

This thesis is based on the publications numbered from I-VI (see below, Section 7.5.1). These publications comprise original research described in the previous chapters of this thesis (Chapters 3 to 6) and a critical review of the literature on EEG and fMRI integration (related to Chapter 1).

7.5 Original contributions

The original contributions of this thesis are summarised as follows (for corresponding publications and software see following sections):

- Chapter 3 (Publication I) provides the first study where different models of how BOLD relates to the frequency profile of neuronal activity have been explicitly compared using human brain imaging data. Results from this study suggest that changes in BOLD are indeed associated with changes in the spectral profile of neuronal activity and that these changes do not arise from one specific

spectral band. Instead they result from the dynamics of the various frequency components together, in particular, from the relative power between high and low frequencies.

- Chapter 4 (Publication II) provides a new Bayesian technique for constructing posterior probability maps (PPMs) for model selection inference at the group level. This approach allows neuroimagers to make inferences about regionally specific effects in the brain, without being restricted to comparing nested models. This method has been used by colleagues to make inferences about how the brain integrates information over time (Publication VI).

- Chapter 5 (Publication IV) provides the first study of how BOLD relates to synaptic and spiking activity in the healthy human brain. Results show that the BOLD signal is dependent upon both synaptic and spiking activities but that the relative contributions of these two inputs are dependent upon the underlying neuronal firing rate. When the underlying neuronal firing is low then the BOLD response is best explained by synaptic activity. However, when the neuronal firing rate is high then both synaptic and spiking activity are required to explain the BOLD signal. In addition, this chapter also provides an efficient modelling framework that can be used to test additional coupling models using EEG/MEG and fMRI data.

- Chapter 6 (Publication V) provides the first application and validation of a recently proposed post-hoc Bayesian model selection approach, using deterministic DCMs for fMRI. This chapter shows that the post-hoc approach provides valid model selection results with the advantage of a substantial reduction in computational time, by decreasing the number of models that need to be inverted from all models to a single model.

7.5.1 Publications

Below are the publications that have arisen from the work developed in this thesis:

I **M. J. Rosa**, J. Kilner, F. Blankenburg, O. Josephs, and W. D. Penny. Estimating the transfer function from neuronal activity to BOLD using simultaneous EEG-fMRI. NeuroImage, 49(2):1496-1509, 2010.

II **M. J. Rosa**, S. Bestmann, L. Harrison, and W. D. Penny. Bayesian Model Selection Maps for Group Studies. NeuroImage, 49(1):217-224, 2010.

III **M. J. Rosa**[1], J. Daunizeau, and K.J. Friston. EEG-fMRI integration : a critical review of biophysical modelling and data analysis approaches. J. Integrative Neurosci., 9(4):453-476, 2010.

IV **M. J. Rosa**, J. Kilner, and W. D. Penny. Bayesian Comparison of Neurovascular Coupling Models using EEG-fMRI. PLoS Computational Biology, 7(6):e1002070, 2011.

V **M. J. Rosa**, K. J. Friston and W. D. Penny. Post-hoc selection of Dynamic Causal Models. Submitted.

VI L. M. Harrison, S. Bestmann, **M. J. Rosa**, W. Penny, and G. G. Green. Time scales of representation in the human brain: weighing past information to predict future events. Frontiers in Human Neuroscience, 5:37, 2011.

7.5.2 Software

Part of the software developed for this thesis has been implemented in the SPM8 version of the Statistical Parametric Mapping software toolbox, which can be downloaded from here: http://www.fil.ion.ucl.ac.uk/spm/. This includes the following functions:

- Chapter 4: *spm_cfg_bms*, *spm_run_bms_map*, *spm_run_bms_disp*, *spm_bms_display*, *spm_bms_display_ROI*, *spm_bms_display_vox*, *spm_bms_partition*, *spm_bms_compare_groups*;
- Chapter 6: *spm_log_evidence_SD*.

The rest of the software developed for this thesis can be downloaded from: http://www.fil.ion.ucl.ac.uk/∼mjoao/archive.html.

[1] Joint first author with J. Daunizeau.

Appendix A

.1 'Heuristic' model of neurovascular coupling

This section summarises the equations of the 'Heuristic' model developed by Kilner et al. [2005]. The full derivation can be found in the original publication [Kilner et al., 2005].

The model proposed by Kilner et al. [2005] assumes that the BOLD signal, b, at any point in time, is proportional to the rate of energy dissipation induced by neuronal transmembrane currents. This assumption is valid under the view that BOLD is thought to be driven by glutamate release and this measure is correlated with energy usage. This relation can be expressed as:

$$b \propto < v^T i > = c < v^T \dot{v} >, \tag{1}$$

where v is the transmembrane potential, i the transmembrane current, and $i = -c\dot{v}$.

Kilner et al. [2005] adopt a simple model of neuronal dynamics of the form:

$$\dot{v_k} = -v_k/\tau_k + u_k, \tag{2}$$

where k represents the neuronal compartment or unit k within a voxel, with time constant τ_k. Kilner et al. [2005] then express Eq. 1 for the BOLD signal as a function of the Jacobian, J, of the neuronal system:

$$b \propto c\mathrm{tr}(< JCov\{v\} >), \tag{3}$$

where $v(t) = \exp^{-Jt} v_0$. J summarises the functional architecture of the system. Each entry of the Jacobian corresponds to either a self-inhibitory connection, $J_{kk} = 1/\tau_k$, or to the intrinsic coupling among units, $J_{kj} = \partial \dot{v}_k / \partial v_j$. Eq. 3 means that the BOLD response is proportional to the trace of the product of the Jacobian (i.e., effective connectivity matrix) and the temporal covariance of the transmembrane potentials.

Kilner et al. [2005] then represent 'neuronal activation' by a variable α, which quantifies changes in effective connectivity and consequently in the dynamics of the neuronal system. This model assumes that the effect of activation, α, is to accelerate the dynamics and increase the system's energy dissipation:

$$\frac{\tilde{b}}{b} \propto \frac{tr(< \tilde{J}Cov\{v\} >)}{tr(< JCov\{v\} >)} = (1 + \alpha), \tag{4}$$

where $\tilde{b}$ corresponds to the BOLD signal during activation. $\tilde{J} = J(\alpha) = (1+\alpha)J$, which is obtained by expanding $J(\alpha)$ around 0 and assuming that $\partial J / \partial \alpha = J(0)$ (the change in intrinsic coupling induced by activation is proportional to the coupling in the resting state, when $\alpha = 0$).

Eq. 4 relates activation to increases in metabolic demands that are then associated with increases in the BOLD signal. Kilner et al. [2005] then relate this result to the effects of the same activation in the EEG signal.

From the point of view of the EEG, activation will cause an acceleration of its dynamics:

$$\tilde{v}(t) = \exp^{-\tilde{J}t} v_0 = v((1 + \alpha)t). \tag{5}$$

This leads to a rougher looking signal with loss of lower frequencies, relative to higher frequencies. A simple way to measure this effect is in terms of the roughness r, which is equivalent to the negative curvature of the EEG autocorrelation function, $\rho(h)$, evaluated at zero lag ($h = 0$):

$$r = -\rho(0)^{''}, \tag{6}$$

where $\rho(h) =< v(t)^T l^T l v(t+h) >$, and l is a lead-field vector. From Eq. 5, the autocorrelation of the signal, $\tilde{\rho}(h)$ can be written as:

$$\begin{aligned} \tilde{\rho}(h) &= \rho((1+\alpha)h) \\ \tilde{\rho}(h)'' &= (1+\alpha)^2 \rho(h)'', \end{aligned} \tag{7}$$

which means that changes in r are related to neuronal activation by:

$$\frac{\tilde{r}}{r} = \frac{\tilde{\rho}(0)''}{\rho(0)''} = (1+\alpha)^2. \tag{8}$$

Knowing that the spectral density, $g(\omega)$, of a random process is the Fourier transform of its autocorrelation, the 'roughness' can be expressed in the frequency domain as:

$$r = \frac{\int \omega^2 g(\omega) d\omega}{\int g(\omega) d\omega}. \tag{9}$$

From Eq. 8 we then obtain the effect of activation on this spectral density:

$$\tilde{g}(\omega) = \frac{g(\omega/(1+\alpha))}{(1+\alpha)}. \tag{10}$$

Eq. 22 shows that the effect of activation in the signal is to shift the spectral profile toward higher frequencies with a reduction in amplitude. This activation can be expressed in terms of the normalised spectral density, $p(\omega)$, as follows:

$$\frac{\tilde{r}}{r} = \frac{\int \omega^2 \tilde{p}(\omega) d\omega}{\int \omega^2 p(\omega) d\omega} = (1+\alpha)^2, \tag{11}$$

where $p(\omega) = g(\omega)/(\int g(\omega) d\omega)$.

Finally, putting Eq. 4 and Eq. 11 together, we obtain the following result:

$$\frac{\tilde{b}}{b} \propto (1+\alpha) \propto \sqrt{\frac{\int \omega^2 \tilde{p}(\omega) d\omega}{\int \omega^2 p(\omega) d\omega}}. \tag{12}$$

The previous equation means that as neuronal activation increases, there is a consequent increase in BOLD signal and a shift in the spectral profile of EEG to higher frequencies. High-frequency dynamics are associated with small effective membrane time constants. These fast changes in potential are responsible for increases in energy consumption to which the BOLD signal is sensitive. The predicted BOLD signal is therefore a function of the frequency profile of neuronal activity as opposed to any particular frequency. For example, an increase in alpha (low-frequency), without a change in total power, would reduce the mean square frequency of EEG and relate to deactivation. Conversely, an increase in gamma (high-frequency) would increase the mean square frequency of EEG signals and relate to activation (Figure 1).

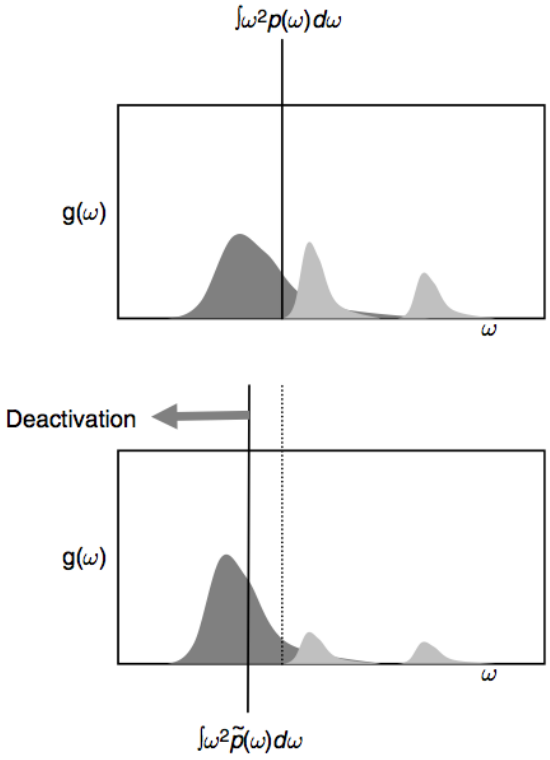

Figure 1: Schematic showing the effect of deactivation on mean square frequency (taken from Kilner et al. [2005]).

.2 Electrovascular model

.2.1 Neural mass model

This section summarises the equations of the neural mass model, which is part of the electrovascular coupling model, proposed by Riera et al. [2006], and used for generating the electrical dynamics in a cortical unit (see Chapter 5 and Riera et al. [2006] for the full derivation). The meaning of each state variable is presented in Table 1.

$$\begin{aligned} \tau_m \frac{dV_{IN}^T(t)}{dt} + V_{IN}^T(t) &= I_3^+(t) R_m^0 \\ \tau_m \frac{dV_{IN}^F(t)}{dt} + V_{IN}^F(t) &= I^+(t) R_m^0 \end{aligned} \tag{13}$$

$$\begin{aligned} \tau_m \frac{d\nu_1(t)}{dt} + \nu_1(t) &= R_m^1 I_1^+(t) \\ \tau_m \frac{d\nu_2(t)}{dt} + \nu_2(t) &= R_m^2 I_2^+(t) \\ \tau_m \frac{d\nu_-(t)}{dt} + \nu_-(t) &= R_m I^-(t) \end{aligned} \tag{14}$$

$$\begin{aligned} \tau_m \frac{dV_{PC}(t)}{dt} &= -(\alpha_0 + \sum_k \frac{1}{\beta_k}) V_{PC}(t) + \frac{\Omega(t)}{\prod_k \beta_k} + R_m I^-(t) \\ &+ \sum_k \left[\frac{R_m \nu_k(t)}{(R_i^k + R_e^k)} - \frac{\nu_-(t)}{\beta_k} \right] \end{aligned} \tag{15}$$

$$\tau_m \frac{d\Omega(t)}{dt} = R_m \sum_k \frac{\beta_k (V_{PC}(t) - \nu_k(t))}{(R_i^k + R_e^k)} + V_{PC}(t) + \nu_-(t) - \Omega(t) \tag{16}$$

$$\begin{aligned}\tau_m \frac{d\rho(t)}{dt} &= -(\alpha_0 + \sum_k \frac{1}{\beta_k})\rho(t) + \frac{\Theta(t)}{\prod_k \beta_k} \\ &+ \frac{R_e^2}{(R_i^2 + R_e^2)}(R_m I^-(t) + R_m^2 I_2^+(t)) \\ &+ \frac{[R_m^1(\nu_-(t) + \nu_2(t)) + R_m(\nu_2(t) - \nu_1(t))]R_e^2}{\prod_k (R_i^k + R_e^k)} \end{aligned} \tag{17}$$

$$\tau_m \frac{d\Theta(t)}{dt} = \left[1 + R_m \sum_k \frac{1}{R_m^k}\right] \rho(t) - \Theta(t). \tag{18}$$

.2.2 Model states and parameters

Tables 1 and 2 summarise the states and parameters of the electrovascular coupling model [Riera et al., 2006] used for generating the electrical and vascular dynamics in a cortical unit (see Chapter 5).

Table 1: Electrical and vascular states.

Type	State	Symbol	Initial value
Electrical			
	Membrane potential at the soma of GABAergic IN (Transmission)	V_{IN}^T	0
	Membrane potential at the soma of GABAergic IN (Feedback)	V_{IN}^F	0
	Membrane potential at the soma of Layer V PC	V_{PC}	0
	Voltage difference	Ω	0
	Equivalent voltage source at the layer V PC basal dendrites	ν_1	0
	Equivalent voltage source at the layer V PC apical tuft dendrites	ν_2	0
	Equivalent voltage source at the soma of layer V PC	ν_-	0
	Extracellular voltage difference		

Table 1 – continued from previous page

	along the layer V PC apical tuft	ρ	0
	Voltage difference	Θ	0
	Time derivative of the input (Balloon approach)	r	0
	Input (Balloon approach)	z	z_0
Vascular			
	Flow-inducing signal	s	0
	CBF	f	1
	CBV	ν	1
	Concentration of dHb	q	1

Table 2: Fixed parameters: these parameters are not estimated from the data.

Type	Symbol	Parameter	Dimension	Value
Electrical				
	c_m	Membrane capacitance	μF/cm^2	0.75
	τ_m	Membrane time constant	ms	30
		Effective membrane resistance:		
	C_m^0	soma compartment (IN)	nF	6.81
	C_m	soma compartment (PC)	nF	1.045
		Membrane resistance:		
	R_m^0	soma compartment (IN)	$G\Omega$	4.082
	R_m	soma compartment (PC)	$G\Omega$	2.871
	R_m^1	basal compartment	$G\Omega$	0.222
	R_m^2	apical tuft compartment	$G\Omega$	0.667
	R_i^1	basal longitudinal (intracellular)	$G\Omega$	0.226
	R_e^1	basal longitudinal (extracellular)	$G\Omega$	0.272
	R_i^2	apical longitudinal (intracellular)	$G\Omega$	2.264
	R_e^2	apical longitudinal (extracellular)	$G\Omega$	2.716
	$\alpha_{_{PC}}$	Layer V PC synaptic factor	$pA(0 \sim 1.5)$	0.4
	$\gamma_{_{IN}}$	GABAergic IN V-A relationships	mV^{-1}	5
	V_0^{IN}	GABAergic IN V-A relationships	mV	0.7
	β_1	Basal voltage divisor	no dim.	2.24
	β_2	Apical voltage divisor	no dim.	7.45
	α_0	Mixed coefficient	no dim.	7.32

Table 2 – continued from previous page

Others				
	χ_{PC}	Layer V PC energetic factor	$nM(0 \sim 1)$	1.0
	χ_{IN}	GABAergic IN energetic factor	$nM(0 \sim 1)$	0.8
	ω_0	Low-pass filter: angular high cut Hz	Hz	$2\pi 8$
	δ	Low-pass filter: damping factor	no dim.	0.8
	A	Low-pass filter: gain	nM^{-1}	1.0
	ρ_{PC}	Layer V PC nonlinear function	no dim.	1.0
	ρ_{IN}	GABAergic IN nonlinear function	no dim.	1.0
	w_{PC}	Layer V PC nonlinear function	$(nA)^2$	0.1091
	w_{IN}	GABAergic IN nonlinear function	$(nA)^2$	0.0464

.2.3 Model comparisons

Table 3 presents the results obtained from comparing all coupling models for all subjects and frequency regimes, as described in Chapter 5.

Table 3: **Model comparison results.** Log-evidence values for each model, session and subject (low-frequencies, high-frequencies and all-frequencies). 'Sum' is the sum of the log-evidences per session and 'Group' the sum of all subjects' sums. The values in bold correspond to the winning model for each session and subject. BF_{12} are the log-Bayes factors between the best (1) and second best (2) model for each session. The values x^* correspond to a posterior probability for the best model higher than 0.95. The values x^{**} correspond to a posterior probability higher than 0.99. 'Freqs.' means the frequencies of the stimuli. $^\diamond$One of the subjects lacked the 4.0 and 7.5 Hz stimulus epochs and was therefore not included in 'low-frequency' analyses. F_x: Log-evidences; B_{xy}: Bayes Factors; ω_{in}: 'mixture' model coefficient (synaptic).

Subject	Session	F_{mix}	F_{in}	F_{out}	BF_{12}	ω_{in}
Low	freq.$^\diamond$	(4-15 Hz)				
1	1	-57.27	-55.69	-56.82	1.13	0.72
	2	-37.27	-34.89	-44.58	2.38	0.80
	3	-59.23	-59.01	-58.21	0.80	0.60
	Sum	-153.77	**-149.59**	-159.61	4.18*	0.71
2	1	-39.54	-38.53	-42.25	1.01	0.66
	2	-41.30	-40.16	-45.85	1.14	0.65
	3	-34.21	-33.40	-37.69	0.81	0.67
	Sum	-115.05	**-112.09**	-125.79	3.00*	0.69

Table 3 – continued from previous page

Group		-268.82	**-261.68**	-285.40	7.14**	0.69
High	frequencies	(12-30 Hz)				
1	1	-37.85	-80.68	-43.22	5.37**	0.60
	2	-32.59	-45.90	-35.21	2.62	0.40
	3	-47.59	-44.56	-87.01	3.03*	0.83
	Sum	**-118.03**	-171.14	-165.44	53.11**	0.61
2	1	-69.48	-69.92	-68.57	0.91	0.55
	2	-47.24	-55.77	-84.73	8.23**	0.48
	3	-59.41	-68.00	-57.21	2.20	0.46
	Sum	**-176.13**	-193.69	-210.51	17.56**	0.50
3	1	-31.91	-43.06	-34.27	2.36	0.46
	2	-38.51	-49.64	-40.64	2.13	0.41
	3	-29.60	-44.52	-31.27	1.67	0.43
	Sum	**-99.72**	-137.22	-106.18	6.46**	0.43
Group		**-393.88**	-502.05	-482.13	88.25**	0.51
All	frequencies	(4-30 Hz)				
1	1	-37.85	-80.68	-43.22	5.37**	0.60
	2	-32.59	-45.90	-35.21	2.62	0.40
	3	-47.59	-44.56	-87.01	3.30*	0.83
	Sum	**-118.03**	-171.14	-165.44	53.11**	0.61
2	1	-89.78	-90.93	-88.10	1.68	0.56
	2	-71.94	-71.42	-71.56	0.14	0.65
	3	-79.46	-91.57	-80.48	1.02	0.59
	Sum	-241.18	-253.92	**-240.14**	1.04	0.60
3	1	-49.41	-61.05	-52.21	2.80	0.49
	2	-54.19	-64.21	-59.13	4.94*	0.46
	3	-44.94	-46.84	-50.76	1.90	0.58
	Sum	**-148.54**	-172.12	-162.10	13.56**	0.51
Group		**-507.75**	-597.18	-567.68	59.93**	0.57

.3 Dynamic causal modelling priors

This thesis uses dynamic causal modelling (DCM) priors in Chapter 6 from 'DCM10', as implemented in SPM8 software (revision 4010). The priors on both the connec-

tivity and hemodynamic parameters are assumed to be Gaussian and factorise over parameter types:

$$p(\theta|m) = p(A|m)p(B|m)p(C|m)p(h|m). \tag{19}$$

The priors on the fixed parameters (A) depend on the number of regions, n, to encourage stable dynamics. The priors on the fixed self-connections (A_{ii}) are defined as follows:

$$p(A_{ii}|m) = N(-1/2, \sigma_{ii}^2), \tag{20}$$

where $\sigma_{ii}^2 = 1/(8 \times n)$. In our case $n = 3$ regions, therefore $\sigma_{ii} = 0.0417$. The priors on the rest of the fixed parameters (A_{ij}) are calculated as follows:

$$p(A_{ij}|m) = N(1/(64 * n), 8/n + 1/(8 \times n)). \tag{21}$$

In our case, this yields $p(A_{ij}|m) = N(0.0052, 2.7083)$. The rest of the connectivity parameters (modulatory and input parameters) have shrinkage priors:

$$\begin{aligned} p(B_{ij}^k|m) &= N(0,1), \\ p(C_{ij}|m) &= N(0,1). \end{aligned} \tag{22}$$

The unknown hemodynamic parameters are $\{\tau_s, \tau_0\}$. These are represented as

$$\begin{aligned} \tau_s &= 0.64 \exp(\theta_{\tau_s}) \\ \tau_0 &= 2 \exp(\theta_{\tau_0}), \end{aligned} \tag{23}$$

and have Gaussian priors:

$$\begin{aligned} p(\theta_{\tau_s}) &= \mathsf{N}(\theta_{\tau_s}; 0, 0.135) \\ p(\theta_{\tau_0}) &= \mathsf{N}(\theta_{\tau_0}; 0, 0.135), \end{aligned} \tag{24}$$

where $h = \{\theta_{\tau_s}, \theta_{\tau_0}\}$ are the hemodynamic parameters to be estimated.
The overall prior density can then be written as $p(\theta|m) = N(\eta, \Sigma)$, where η and

Σ are concatenations of the above means and variances.

References

Robert J. Adler. Random fields. In *Encyclopedia of Environmetrics*. John Wiley and Sons, Ltd, 2006. ISBN 9780470057339. doi: 10.1002/9780470057339.var007. 14, 16

H Akaike. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory*, volume 1, pages 267–281. Akademiai Kiado, 1973. 42

P. Allen, O. Josephs, and R. Turner. A method for removing imaging artifact from continuous EEG recorded during functional MRI. *Neuroimage*, 12(2):230–239, 2000. 7, 54

Philip J. Allen, Giovanni Polizzi, Karsten Krakow, David R. Fish, and Louis Lemieux. Identification of eeg events in the mr scanner: The problem of pulse artifact and a method for its subtraction. *NeuroImage*, 8(3):229 – 239, 1998. 7

J. L. Andersson, C. Hutton, J. Ashburner, R. Turner, and K. Friston. Modeling geometric deformations in EPI time series. *Neuroimage*, 13:903–919, May 2001. 84

Anonymous. Connecting the dots. *Nat. Neurosci.*, 12:99, Feb 2009. ISSN 0006–3495. 99

J. S. Archer, D. F. Abbott, A. B. Waites, and G. D. Jackson. fMRI "deactivation" of the posterior cingulate during generalized spike and wave. *Neuroimage*, 20: 1915–1922, Dec 2003. 79

J. Ashburner and K.J. Friston. Voxel-based morphometry – the methods. *NeuroImage*, 11:805–821, 2000. 16

A. Aubert and R. Costalat. Interaction between astrocytes and neurons studied using a mathematical model of compartmentalized energy metabolism. *J. Cereb. Blood Flow Metab.*, 25:1476–1490, Nov 2005. 100

A. Babajani and H. Soltanian-Zadeh. Integrated MEG/EEG and fMRI model based on neural masses. *Biomedical Engineering, IEEE Transactions on*, 53(9):1794–1801, September 2006. 50, 101

C. Babiloni, V. Pizzella, C. D. Gratta, A. Ferretti, and G. L. Romani. Fundamentals of electroencephalography, magnetoencephalography, and functional magnetic resonance imaging. *Int. Rev. Neurobiol.*, 86:67–80, 2009. 107

S. Baillet, J. C. Mosher, and R. M. Leahy. Electromagnetic brain mapping. *IEEE signal processing magazine*, 18:14–30, November 2001a. 5, 100, 112

S. Baillet, J. J. Riera, G. Marin, J. F. Mangin, J. Aubert, and L. Garnero. Evaluation of inverse methods and head models for EEG source localization using a human skull phantom. *Phys Med Biol*, 46:77–96, Jan 2001b. 5

M. Beal and Z. Ghahramani. The Variational Bayesian EM algorithms for incomplete data: with application to scoring graphical model structures. In J. Bernardo, M. Bayarri, J. Berger, and A. Dawid, editors, *Bayesian Statistics 7*. Cambridge University Press, 2003. 41, 42, 43, 131

T. E. Behrens, L. T. Hunt, M. W. Woolrich, and M. F. Rushworth. Associative learning of social value. *Nature*, 456:245–249, Nov 2008. 82

J.M. Bernardo and A.M. Smith. Bayesian Theory. *Meas. Sci. Technol.*, 12:221–222, 2001. 45

S. Bestmann, L. M. Harrison, F. Blankenburg, R. B. Mars, P. Haggard, K. J. Friston, and J. C. Rothwell. Influence of uncertainty and surprise on human corticospinal excitability during preparation for action. *Curr. Biol.*, 18:775–780, May 2008. 82, 84, 85

C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2nd edition, 2006. 36, 41

G. Bonmassar, D. P. Schwartz, A. K. Liu, K. K. Kwong, A. M. Dale, and J. W. Belliveau. Spatiotemporal brain imaging of visual-evoked activity using interleaved eeg and fmri recordings. *NeuroImage*, 13(6):1035 – 1043, 2001. 8

G. Bonvento, N. Sibson, and L. Pellerin. Does glutamate image your thoughts? *Trends Neurosci.*, 25:359–364, Jul 2002. 10, 100

L. Boorman, A. J. Kennerley, D. Johnston, M. Jones, Y. Zheng, P. Redgrave, and J. Berwick. Negative blood oxygen level dependence in the rat: a model for investigating the role of suppression in neurovascular coupling. *J. Neurosci.*, 30: 4285–4294, Mar 2010. 28, 162

G. M. Boynton, S. A. Engel, G. H. Glover, and D. J. Heeger. Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.*, 16:4207–4221, Jul 1996. 14

P. Brown and J. F. Marsden. Cortical network resonance and motor activity in humans. *Neuroscientist*, 7:518–527, Dec 2001. 22

C. Buchel and K. J. Friston. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. *Cereb. Cortex*, 7:768–778, Dec 1997. 132, 141, 151

C. Buchel, R. J. Wise, C. J. Mummery, J. B. Poline, and K. J. Friston. Nonlinear regression in parametric activation studies. *Neuroimage*, 4:60–66, Aug 1996. 84

D. G. Buerk, B. M. Ances, J. H. Greenberg, and J. A. Detre. Temporal dynamics of brain tissue nitric oxide during functional forepaw stimulation in rats. *Neuroimage*, 18:1–9, Jan 2003. 108

G. R. Burkitt, R. B. Silberstein, P. J. Cadusch, and A. W. Wood. Steady-state visual evoked potentials and travelling waves. *Clin Neurophysiol*, 111:246–258, Feb 2000. 63

R. B. Buxton, E. C. Wong, and L. R. Frank. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn Reson Med*, 39:855–864, Jun 1998. 20, 100, 105, 134

G. Carmignoto and M. Gomez-Gonzalo. The contribution of astrocyte signalling to neurovascular coupling. *Brain Res Rev*, 63:138–148, May 2010. 11, 100

R. Christensen. *Plane Answers to Complex Questions: The Theory of Linear Models*. Springer, 3rd edition, 2002. 14, 61

J. Chumbley, K.J. Friston, T. Fearn, and S.J. Kiebel. A Metropolis-Hastings algorithm for dynamic causal models. *NeuroImage*, 38(3):478–487, 2007. doi: 10.1016/j.neuroimage.2007.07.028. 39

Anders M. Dale, Arthur K. Liu, Bruce R. Fischl, Randy L. Buckner, John W. Belliveau, Jeffrey D. Lewine, and Eric Halgren. Dynamic statistical parametric mapping: Combining fmri and meg for high-resolution imaging of cortical activity. *Neuron*, 26(1):55 – 67, 2000. 9

J. Daunizeau, K. J. Friston, and S. J. Kiebel. Variational Bayesian identification and prediction of stochastic nonlinear dynamic causal models. *Physica D*, 238: 2089–2118, Nov 2009. 128

Jean Daunizeau, Christophe Grova, Guillaume Marrelec, Jrmie Mattout, Saad Jbabdi, Mlanie Plgrini-Issac, Jean-Marc Lina, and Habib Benali. Symmetrical event-related eeg/fmri information fusion in a variational bayesian framework. *NeuroImage*, 36(1):69 – 87, 2007. 9

O. David and K. J. Friston. A neural mass model for MEG/EEG: coupling and neuronal dynamics. *Neuroimage*, 20:1743–1755, Nov 2003. 20, 100

P. Dayan and L. F. Abbott. *Theoretical Neuroscience*. MIT Press, 1st edition, 2002. 12

S. Debener, M. Ullsperger, M. Siegel, and A. K. Engel. Single-trial EEG-fMRI reveals the dynamics of cognitive function. *Trends Cogn. Sci. (Regul. Ed.)*, 10:558–563, Dec 2006. 7

Gustavo Deco, Viktor K. Jirsa, Peter A. Robinson, Michael Breakspear, and Karl Friston. The dynamic brain: From spiking neurons to neural masses and cortical fields. *PLoS Comput Biol*, 4(8):e1000092, 08 2008. 20

R. Deichmann, C. Schwarzbauer, and R. Turner. Optimisation of the 3D MDEFT sequence for anatomical brain imaging: technical implications at 1.5 and 3 T. *Neuroimage*, 21:757–767, Feb 2004. 54

H. E. den Ouden, K. J. Friston, N. D. Daw, A. R. McIntosh, and K. E. Stephan. A dual role for prediction error in associative learning. *Cereb. Cortex*, 19:1175–1185, May 2009. 26

James M. Dickey. The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Mathematical Statistics*, 42(1):pp. 204–223, 1971. ISSN 00034851. 43, 132, 135, 137

T. Eichele, K. Specht, M. Moosmann, M. L. Jongsma, R. Q. Quiroga, H. Nordby, and K. Hugdahl. Assessing the spatiotemporal evolution of neuronal activation with single-trial event-related potentials and functional MRI. *Proc. Natl. Acad. Sci. U.S.A.*, 102:17798–17803, Dec 2005. 52, 160

T. Eichele, V. D. Calhoun, and S. Debener. Mining EEG-fMRI using independent component analysis. *Int J Psychophysiol*, Feb 2009. 52, 160

S. B. Eickhoff, K. E. Stephan, H. Mohlberg, C. Grefkes, G. R. Fink, K. Amunts, and K. Zilles. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage*, 25:1325–1335, May 2005. 67

A. Ekstrom, N. Suthana, D. Millett, I. Fried, and S. Bookheimer. Correlation between BOLD fMRI and theta-band local field potentials in the human hippocampal area. *J. Neurophysiol.*, 101:2668–2678, May 2009. 9, 11, 30

G. B. Ermentrout and D Terman. *Mathematical Foundations of Neuroscience*. Springer, 1st edition, 2010. 20

C. Estrada and J. DeFelipe. Nitric oxide-producing neurons in the neocortex: morphological and functional relationship with intraparenchymal microvasculature. *Cereb. Cortex*, 8:193–203, 1998. 10, 100, 108, 129

S. L. Fairhall and A. Ishai. Effective connectivity within the distributed cortical network for face perception. *Cereb. Cortex*, 17:2400–2406, Oct 2007. 26

Jessica A. Filosa and Vctor M. Blanco. Neurovascular coupling in the mammalian brain. *Experimental Physiology*, 92(4):641–646, 2007. 11

R.S.J. Frackowiak, K.J. Friston, C. Frith, R. Dolan, C.J. Price, S. Zeki, J. Ashburner, and W.D. Penny. *Human Brain Function*. Academic Press, 2nd edition, 2003. 49

K. Friston. Functional integration and inference in the brain. *Prog. Neurobiol.*, 68: 113–143, Oct 2002a. 22

K. Friston and W. Penny. Post hoc Bayesian model selection. *Neuroimage*, 56: 2089–2099, Jun 2011. 43, 132, 135, 136, 140, 141, 154, 155

K. Friston, J. Mattout, N. Trujillo-Barreto, J. Ashburner, and W. Penny. Variational free energy and the Laplace approximation. *Neuroimage*, 34:220–234, Jan 2007a. 23, 37, 38, 40, 42, 102, 131, 135, 154

K. J. Friston. Bayesian estimation of dynamical systems: an application to fMRI. *Neuroimage*, 16:513–530, Jun 2002b. xii, 19, 20, 21, 107

K. J. Friston. Neurophysiology: The Brain at Work. *Current Biology*, 18:418–420, May 2008. 10, 12, 50, 100

K. J. Friston, P. Rotshtein, J. J. Geng, P. Sterzer, and R. N. Henson. A critique of functional localisers. *Neuroimage*, 30:1077–1087, May 2006. 87

K. J. Friston, N. Trujillo-Barreto, and J. Daunizeau. DEM: a variational treatment of dynamic systems. *Neuroimage*, 41:849–885, Jul 2008. 128

K. J. Friston, B. Li, J. Daunizeau, and K. E. Stephan. Network discovery with DCM. *Neuroimage*, Dec 2010a. 23, 132, 133, 142

K.J. Friston. Functional integration. In R.S.J. Frackowiak, K.J. Friston, C. Frith, R. Dolan, K.J. Friston, C.J. Price, S. Zeki, J. Ashburner, and W.D. Penny, editors, *Human Brain Function*. Academic Press, 2nd edition, 2003a. 12

K.J. Friston. Dynamic causal models. In R.S.J. Frackowiak, K.J. Friston, C. Frith, R. Dolan, K.J. Friston, C.J. Price, S. Zeki, J. Ashburner, and W.D. Penny, editors, *Human Brain Function*. Academic Press, 2nd edition, 2003b. 25

K.J. Friston and W.D. Penny. Posterior probability maps and SPMs. *NeuroImage*, 19(3):1240–1249, 2003. xi, 17, 18, 35, 37, 86, 97, 131, 133

K.J. Friston, A.P. Holmes, K.J. Worsley, J.B. Poline, C. Frith, and R.S.J. Frackowiak. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2:189–210, 1995. 14, 50, 60

K.J. Friston, C. Buchel, G.R. Fink, J. Morris, E. Rolls, and R. Dolan. Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, 6:218–229, 1997. doi: 10.1006/nimg.1997.0291. 141

K.J. Friston, A. Mechelli, R. Turner, and C.J. Price. Nonlinear responses in fMRI: The Balloon model, Volterra kernels and other hemodynamics. *NeuroImage*, 12: 466–477, 2000. 20, 100, 105, 107, 134

K.J. Friston, D.E. Glaser, R.N.A. Henson, S.J. Kiebel, C. Phillips, and J. Ashburner. Classical and Bayesian inference in neuroimaging: Applications. *NeuroImage*, 16: 484–512, 2002a. 86

K.J. Friston, W.D. Penny, C. Phillips, S.J. Kiebel, G. Hinton, and J. Ashburner. Classical and Bayesian inference in neuroimaging: Theory. *NeuroImage*, 16:465–483, 2002b. 16, 36, 86

K.J. Friston, L. Harrison, and W.D. Penny. Dynamic Causal Modelling. *NeuroImage*, 19(4):1273–1302, 2003. 22, 23

K.J. Friston, R.N.A. Henson, C. Phillips, and J. Mattout. Bayesian estimation of evoked and induced responses. *Human Brain Mapping*, 27:722–735, 2005a. doi: 10.1002/hbm.20214. 102

K.J. Friston, W.D. Penny, and O. David. Modelling brain responses. In M.F. Glabus, editor, *International Review of Neurobiology, Neuroimaging, Part A*. Elsevier, 2005b. 12

K.J. Friston, J. Ashburner, S.J. Kiebel, T.E. Nichols, and W.D. Penny, editors. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press, 2007b. 58, 160

K.J. Friston, K.E. Stephan, Baojuan Li, and J. Daunizeau. Generalised filtering. *Mathematical Problems in Engineering*, 621670, 2010b. doi: 10.1155/2010/621670. 141

R. D. Frostig, E. E. Lieke, D. Y. Ts'o, and A. Grinvald. Cortical functional architecture and local coupling between neuronal activity and the microcirculation revealed by in vivo high-resolution optical imaging of intrinsic signals. *Proc. Natl. Acad. Sci. U.S.A.*, 87:6082–6086, Aug 1990. 2

M. I. Garrido, J. M. Kilner, S. J. Kiebel, K. E. Stephan, T. Baldeweg, and K. J. Friston. Repetition suppression and plasticity in the human brain. *Neuroimage*, 48:269–279, Oct 2009. 26

W. S. Geisler and D. G. Albrecht. Visual cortex neurons in monkeys and cats: detection, discrimination, and identification. *Vis. Neurosci.*, 14:897–919, 1997. 10

A. Gelman, J. Carlin, H. Stern, and D. Rubin, editors. *Bayesian data analysis*. Chapman and Hall, 1995. 17, 38, 97

M. Girolami. Bayesian inference for differential equations. *Theoretical Computer Science*, 408(1):4 – 16, 2008. 38

J.B. Goense and N.K. Logothetis. Neurophysiology of the BOLD fMRI signal in awake monkeys. *Curr. Biol.*, 18:631–640, May 2008. 10, 27, 51, 56, 77

R. Goldman, J. Stern, J. Engel, and M. Cohen. Simultaneous EEG and fMRI of the alpha rhythm. *Neuroreport*, 18(13):2587–2492, 2002. 11, 30, 50, 51, 78, 160

R.I. Goldman, J.M. Stern, J. Engel, and M.S. Cohen. Acquiring simultaneous EEG and functional MRI. *Clin Neurophysiol*, 111:1974–1980, Nov 2000. 54

S.I. Gonçalves, J.C. de Munck, P.J. Pouwels, R. Schoonhoven, J.P. Kuijer, N.M. Maurits, J.M. Hoogduin, E.J. Van Someren, R.M. Heethaar, and F.H. Lopes da Silva. Correlating the alpha rhythm to BOLD using simultaneous EEG/fMRI: inter-subject variability. *Neuroimage*, 30:203–213, Mar 2006. 78

C. Grefkes, D. A. Nowak, S. B. Eickhoff, M. Dafotakis, J. Kust, H. Karbe, and G. R. Fink. Cortical connectivity after subcortical stroke assessed with functional magnetic resonance imaging. *Ann. Neurol.*, 63:236–246, Feb 2008. 26

F. Grimbert and O. Faugeras. Bifurcation analysis of Jansen's neural mass model. *Neural Comput*, 18:3052–3068, Dec 2006. 20

M. J. Grol, J. Majdandzi?, K. E. Stephan, L. Verhagen, H. C. Dijkerman, H. Bekkering, F. A. Verstraten, and I. Toni. Parieto-frontal connectivity during visually guided grasping. *J. Neurosci.*, 27:11877–11887, Oct 2007. 26

M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa. Magnetoencephalography - theory, instrumentation, and applications to noninvasive

studies of the working human brain. *Rev. Mod. Phys.*, 65(2):413–497, Apr 1993. 3, 49

K. Hamandi, A. Salek-Haddadi, D. R. Fish, and L. Lemieux. EEG/functional MRI in epilepsy: The Queen Square Experience. *J Clin Neurophysiol*, 21:241–248, 2004. 79

Sam Harris, Myles Jones, Ying Zheng, and Jason Berwick. Does neural input or processing play a greater role in the magnitude of neuroimaging signals? *Frontiers in Neuroenergetics*, 2(0), 2010. ISSN 1662-6427. 161

L. M. Harrison, A. Duggins, and K. J. Friston. Encoding uncertainty in the hippocampus. *Neural Netw*, 19:535–546, Jun 2006. 85

L. M. Harrison, W. Penny, J. Daunizeau, and K. J. Friston. Diffusion-based spatial priors for functional magnetic resonance images. *Neuroimage*, 41:408–423, Jun 2008. 86

L. M. Harrison, S. Bestmann, M. J. Rosa, W. Penny, and G. G. Green. Time scales of representation in the human brain: weighing past information to predict future events. *Front Hum Neurosci*, 5:37, 2011. 94, 97

N. V. Hartvig and J. L. Jensen. Spatial mixture modeling of fMRI data. *Hum Brain Mapp*, 11:233–248, Dec 2000. 86

D. A. Harville. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72(358): pp. 320–338, 1977. ISSN 01621459. 36

J. D. Haynes, J. Tregellas, and G. Rees. Attentional integration between anatomically distinct stimulus representations in early visual cortex. *Proc. Natl. Acad. Sci. U.S.A.*, 102:14925–14930, Oct 2005. 26

B. J. He, A. Z. Snyder, J. M. Zempel, M. D. Smyth, and M. E. Raichle. Electrophysiological correlates of the brain's intrinsic large-scale functional architecture. *Proc. Natl. Acad. Sci. U.S.A.*, 105:16039–16044, Oct 2008. 31

D. J. Heeger, A. C. Huk, W. S. Geisler, and D. G. Albrecht. Spikes versus BOLD: what does neuroimaging tell us about neuronal activity? *Nat. Neurosci.*, 3:631–633, Jul 2000. 10, 27, 128

W. E. Hick. A note on one-tailed and two-tailed tests. *Psychol Rev*, 59:316–318, Jul 1952.

Jennifer A. Hoeting, David Madigan, Adrian E. Raftery, and Chris T. Volinsky. Bayesian model averaging: A tutorial. *Statistical Science*, 14(4):382–401, 1999. 34

B. Horwitz, K. J. Friston, and J. G. Taylor. Neural modeling and functional brain imaging: an overview. *Neural Netw*, 13:829–846, 2000. 18, 26

S. Huettel, A. Song, and G. McCarthy. *Functional Magnetic Resonance Imaging*. Sinauer Associates, 2nd edition, 2009. 2, 3

R. Hyman. Stimulus information as a determinant of reaction time. *J Exp Psychol*, 45:188–196, Mar 1953. 84

B. H. Jansen and V. G. Rit. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biol Cybern*, 73: 357–366, Sep 1995. 20, 100

P. Jezzard, P. Matthews, and S. Smith. *Functional MRI: An introduction to methods*. Oxford University Press, 1st edition, 2001. 2

R.E. Kass and A.E. Raftery. Bayes factors. *J. Am. Stat. Assoc.*, 90:773 795, 1995. 39, 43

C. Kayser, M. Kim, K. Ugurbil, D. S. Kim, and P. Konig. A comparison of hemodynamic and neural responses in cat visual cortex using complex stimuli. *Cereb. Cortex*, 14:881–891, Aug 2004. 28

S. J. Kiebel, M. I. Garrido, R. Moran, C. C. Chen, and K. J. Friston. Dynamic causal modeling for EEG and MEG. *Hum Brain Mapp*, 30:1866–1876, Jun 2009. 22, 23, 133, 162

Stefan J. Kiebel, Jean Daunizeau, and Karl J. Friston. A hierarchy of Time-Scales and the Brain. *PLoS Comput Biol*, 4(11):e1000209, 11 2008. 98

J. Kilner and K.J. Friston. Topological inference for EEG and MEG. *The Annals of Applied Statistics*, 4(3):1272–1290, 2010. doi: 10.1214/10-AOAS337. 15

J. Kilner, J. Mattout, R. Henson, and K.J. Friston. Hemodynamic correlates of EEG: A heuristic. *NeuroImage*, 28:280–286, 2005. xi, xxi, 8, 51, 56, 78, 79, 101, 166, 167, 169

J. Kilner, J. Niessing, and R. Galuske. Multimodal fusion: Neuronal activity and the hemodynamic response. *Conference Abstract: Human Brain Mapping, Chicago, USA*, 2007. 51

N. Kopell, G.B. Ermentrout, M.A. Whittington, and R.D. Traub. Gamma rhythms and beta rhythms have different synchronization properties. *Proc. Natl. Acad. Sci. U.S.A.*, 97:1867–1872, Feb 2000. 79

K. Krakow, L. Lemieux, D. Messina, C. A. Scott, M. R. Symms, J. S. Duncan, and D. R. Fish. Spatio-temporal imaging of focal interictal epileptiform activity using EEG-triggered functional MRI. *Epileptic Disord*, 3:67–74, Jun 2001. 79

Ruth M. Krebs, Marty G. Woldorff, Claus Tempelmann, Nils Bodammer, Toemme Noesselt, Carsten N. Boehler, Henning Scheich, Jens-Max Hopf, Emrah Duzel, Hans-Jochen Heinze, and Mircea A. Schoenfeld. High-field fmri reveals brain activation patterns underlying saccade execution in the human superior colliculus. *PLoS ONE*, 5(1):e8691, 01 2010. 2

F. Kruggel, C.J. Wiggins, C.S. Herrmann, and D.Y. von Cramon. Recording of the event-related potentials during functional mri at 3.0 tesla field strength. *Magnetic Resonance in Medicine*, 44(2):277–282, 2000. 7

K. K. Kwong, J. W. Belliveau, D. A. Chesler, I. E. Goldberg, R. M. Weisskoff, B. P. Poncelet, D. N. Kennedy, B. E. Hoppel, M. S. Cohen, and R. Turner. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc. Natl. Acad. Sci. U.S.A.*, 89:5675–5679, Jun 1992. 2

J.P. Lachaux, P. Fonlupt, P. Kahane, L. Minotti, D. Hoffmann, O. Bertrand, and M. Baciu. Relationship between task-related gamma oscillations and BOLD signal: new insights from combined fMRI and intracranial EEG. *Hum Brain Mapp*, 28: 1368–1375, Dec 2007. 30, 50

H. Laufs. Endogenous brain oscillations and related networks detected by surface EEG-combined fMRI. *Hum Brain Mapp*, 29:762–769, Jul 2008. 161

H. Laufs, A. Kleinschmidt, A. Beyerle, E. Eger, A. Salek-Haddadi, C. Preibisch, and K. Krakow. EEG-correlated fMRI of human alpha activity. *NeuroImage*, 4(19): 1463–1476, 2003. 11, 50, 51, 55, 160

H. Laufs, J.L. Holt, R. Elfont, M. Krams, J.S. Paul, K. Krakow, and A. Kleinschmidt. Where the BOLD signal goes when alpha EEG leaves. *Neuroimage*, 31:1408–1418, Jul 2006a. 30, 51, 77

H. Laufs, U. Lengler, K. Hamandi, A. Kleinschmidt, and K. Krakow. Linking generalized spike-and-wave discharges and resting state brain activity by using EEG/fMRI in a patient with absence seizures. *Epilepsia*, 47:444–448, Feb 2006b. 52

H. Laufs, J. Daunizeau, D. W. Carmichael, and A. Kleinschmidt. Recent advances in recording electrophysiological data simultaneously with magnetic resonance imaging. *Neuroimage*, 40:515–528, Apr 2008. 7

J. H. Lee, R. Durand, V. Gradinaru, F. Zhang, I. Goshen, D. S. Kim, L. E. Fenno, C. Ramakrishnan, and K. Deisseroth. Global and local fMRI signals driven by neurons defined optogenetically by type and wiring. *Nature*, 465:788–792, Jun 2010. 29, 128, 162

A. P. Leff, T. M. Schofield, K. E. Stephan, J. T. Crinion, K. J. Friston, and C. J. Price. The cortical dynamics of intelligible speech. *J. Neurosci.*, 28:13209–13215, Dec 2008. 26

L. Lemieux, A. Salek-Haddadi, O. Josephs, P. Allen, N. Toms, C. Scott, K. Krakow, R. Turner, and D.R. Fish. Event-related fMRI with simultaneous and continuous EEG: description of the method and initial case report. *Neuroimage*, 14:780–787, Sep 2001. 8, 49, 52, 160

J. Li and C. Iadecola. Nitric oxide and adenosine mediate vasodilation during functional activation in cerebellar cortex. *Neuropharmacology*, 33:1453–1461, Nov 1994. 108

V. Litvak and K.J. Friston. Electromagnetic source reconstruction for group studies. *NeuroImage*, 2008. 97, 161

V. Litvak, J. Mattout, S.J. Kiebel, C. Phillips, R.N.A. Henson, J. Kilner, G. Barnes, R. Oostenveld, J. Daunizeau, G. Flandin, W.D. Penny, and K.J. Friston. EEG and

MEG data analysis in SPM8. *Comput. Intell. Neurosci.*, 2011(852961), 2011. doi: 10.1155/2011/852961. 15

A. K. Liu, J. W. Belliveau, and A. M. Dale. Spatiotemporal imaging of human brain activity using functional MRI constrained magnetoencephalography data: Monte Carlo simulations. *Proc. Natl. Acad. Sci. U.S.A.*, 95:8945–8950, Jul 1998. 9

Z. Liu, F. Kecman, and B. He. Effects of fMRI-EEG mismatches in cortical current density estimation integrating fMRI and EEG: a simulation study. *Clin Neurophysiol*, 117:1610–1622, Jul 2006. 9

N. K. Logothetis and J. Pfeuffer. On the nature of the BOLD fMRI contrast mechanism. *Magn Reson Imaging*, 22:1517–1531, Dec 2004. 3

N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann. Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412:150–157, July 2001. 10, 27, 51, 127

N.K. Logothetis. What we can do and what we cannot do with fMRI. *Nature*, 453: 869–878, Jun 2008. 5, 6, 79, 99

N. J. Maandag, D. Coman, B. G. Sanganahalli, P. Herman, A. J. Smith, H. Blumenfeld, R. G. Shulman, and F. Hyder. Energetics of neuronal signaling and fMRI activity. *Proc. Natl. Acad. Sci. U.S.A.*, 104:20546–20551, Dec 2007. 128

A. Maier, M. Wilke, C. Aura, C. Zhu, F. Q. Ye, and D. A. Leopold. Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. *Nat. Neurosci.*, 11:1193–1200, Oct 2008. 28

H. Mandelkow, P. Halder, P. Boesiger, and D. Brandeis. Synchronization facilitates removal of MRI artifacts from concurrent EEG recordings and increases usable bandwidth. *NeuroImage*, 32(3):1120–1126, 2006. 53

A. C. Marreiros, J. Daunizeau, S. J. Kiebel, and K. J. Friston. Population dynamics: variance and the sigmoid activation function. *Neuroimage*, 42:147–157, Aug 2008a. 100, 133

A. C. Marreiros, S. J. Kiebel, J. Daunizeau, L. M. Harrison, and K. J. Friston. Population dynamics under the Laplace assumption. *Neuroimage*, 44:701–714, Feb 2009. 20

A.C. Marreiros, S.J. Kiebel, and K.J. Friston. Dynamic causal modelling for fMRI: A two-state model. *NeuroImage*, 39:269–278, 2008b. doi: 10.1016/j.neuroimage.2007.08.019. 23, 162

R. B. Mars, S. Debener, T. E. Gladwin, L. M. Harrison, P. Haggard, J. C. Rothwell, and S. Bestmann. Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. *J. Neurosci.*, 28: 12539–12545, Nov 2008. 85

J. J. Marshall, E. P. Wei, and H. A. Kontos. Independent blockade of cerebral vasodilation from acetylcholine and nitric oxide. *Am. J. Physiol.*, 255:H847–854, Oct 1988. 108

A.R. McIntosh and F. Gonzalez-Lima. Cortical network resonance and motor activity in humans. *Human Brain Mapping*, 2:2–22, 1994. 22

K.J. Miller, E.C. Leuthardt, G. Schalk, R.P. Rao, N.R. Anderson, D.W. Moran, J.W. Miller, and J.G. Ojemann. Spectral changes in cortical surface potentials during motor movement. *J. Neurosci.*, 27:2424–2432, Feb 2007. 79

P. R. Montague, S. E. Hyman, and J. D. Cohen. Computational roles for dopamine in behavioural control. *Nature*, 431:760–767, Oct 2004. 82, 97

M. Moosmann, P. Rittera, I. Krastela, A. Brinka, S. Theesa, F. Blankenburg, B. Taskina, H. Obriga, and A. Villringer. Correlates of alpha rhythm in functional magnetic resonance imaging and near infrared spectroscopy. *NeuroImage*, 20(1):145–158, 2003. 11, 30, 50, 52, 55, 160

R. J. Moran, K. E. Stephan, S. J. Kiebel, N. Rombach, W. T. O'Connor, K. J. Murphy, R. B. Reilly, and K. J. Friston. Bayesian estimation of synaptic physiology from the spectral responses of neural masses. *Neuroimage*, 42:272–284, Aug 2008. 102, 107, 129

R. J. Moran, K. E. Stephan, T. Seidenbecher, H. C. Pape, R. J. Dolan, and K. J. Friston. Dynamic causal models of steady-state responses. *Neuroimage*, 44:796–811, Feb 2009. 23, 133

J. Mourao-Miranda, A. L. Bokde, C. Born, H. Hampel, and M. Stetter. Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data. *Neuroimage*, 28:980–995, Dec 2005. 22

R. Mukamel, H. Gelbard, A. Arieli, U. Hasson, I. Fried, and R. Malach. Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. *Science*, 309:951–954, Aug 2005. 11, 30, 51, 77, 128

P. Muller, G. Parmigiani, and K. Rice. FDR and Bayesian Multiple Comparisons Rules. In *Bayesian Statistics 8: Proceedings of the Eighth Valencia International Meeting*, July 2007. 97

S.D. Muthukumaraswamy and K.D. Singh. Functional decoupling of BOLD and gamma-band amplitudes in human primary visual cortex. *Hum Brain Mapp*, Aug 2008. 78

D.G. Nair. About being BOLD. *Brain Res. Brain Res. Rev.*, 50:229–243, Dec 2005. 50

C. Nangini, F. Tam, and S.J. Graham. A novel method for integrating MEG and BOLD fMRI signals with the linear convolution model in human primary somatosensory cortex. *Hum Brain Mapp*, 29:97–106, Jan 2008. 50

Radford M. Neal and Geoffrey E. Hinton. A new view of the EM algorithm that justifies incremental and other variants. In *Learning in Graphical Models*, pages 335–370. Kluwer Academic Publishers, 1996. 39

E. Niedermeyer and F. H. Lopes da Silva. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams and Wilkins, 5th edition, 2004. 3, 4

J. Niessing, B. Ebisch, K.E. Schmidt, M. Niessing, W. Singer, and R.A. Galuske. Hemodynamic signals correlate tightly with synchronized gamma oscillations. *Science*, 309:948–951, Aug 2005. 28, 51

Y. Nir, L. Fisch, R. Mukamel, H. Gelbard-Sagiv, A. Arieli, I. Fried, and R. Malach. Coupling between neuronal firing rate, gamma LFP, and BOLD fMRI is related to interneuronal correlations. *Curr. Biol.*, 17:1275–1285, Aug 2007. 31

U. Noppeney, O. Josephs, J. Hocking, C. J. Price, and K. J. Friston. The effect of prior visual information on recognition of speech and sounds. *Cereb. Cortex*, 18: 598–609, Mar 2008. 26

P. L. Nunez. Localization of brain activity with electroencephalography. *Adv Neurol*, 54:39–65, 1990. 5

P.L. Nunez and B. Cutillo. *Neocortical Dynamics and Human EEG Rhythms*. Oxford University Press, 1st edition, 1995. 4, 5

P.L. Nunez and R.B. Silberstein. On the relationship of synaptic activity to macroscopic measurements: does co-registration of EEG with fMRI make sense? *Brain Topogr*, 13:79–96, 2000. 5, 50

J. P. O'Doherty, A. Hampton, and H. Kim. Model-based fMRI and its application to reward learning and decision making. *Ann. N. Y. Acad. Sci.*, 1104:35–53, May 2007. 97, 158

S. Ogawa, D. W. Tank, R. Menon, J. M. Ellermann, S. G. Kim, H. Merkle, and K. Ugurbil. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc. Natl. Acad. Sci. U.S.A.*, 89:5951–5955, Jul 1992. 2

L. Parkes, P. Fries, C. Kerskens, and D. Norris. Reduced BOLD response to periodic visual stimulation. *NeuroImage*, 21(1):236–243, 2004. 68, 119, 128

L. Pauling and C. Coryell. The magnetic properties and structure of hemoglobin, oxyhemoglobin and carbonmonoxyhemoglobin. *PNAS*, 22(4):210–216, March 1936. 2

W. Penny and A. Holmes. Random effects analysis. In K. Friston, J. Ashburner, S. Kiebel, T. Nichols, and W. Penny, editors, *Statistical Parametric Mapping: The analysis of functional brain images*. Elsevier, London, 2006. 53, 62, 73, 81

W. D. Penny. Comparing dynamic causal models using AIC, BIC and free energy. *Neuroimage*, 59:319–330, Jan 2012. 41, 42, 43, 131, 135

W. D. Penny, K. E. Stephan, A. Mechelli, and K. J. Friston. Comparing dynamic causal models. *Neuroimage*, 22:1157–1172, Jul 2004. 25, 102, 123, 126

W.D. Penny, S.J. Kiebel, and K.J. Friston. Variational Bayesian Inference for fMRI time series. *NeuroImage*, 19(3):727–741, 2003. 40

W.D. Penny, N. Trujillo-Bareto, and K.J. Friston. Bayesian fMRI time series analysis with spatial priors. *NeuroImage*, 24(2):350–362, 2005. 35, 86

W.D. Penny, G. Flandin, and N. Trujillo-Barreto. Bayesian Comparison of Spatially Regularised General Linear Models. *Human Brain Mapping*, 28(4):275–293, 2007. 82, 86, 160

W.D. Penny, K.E. Stephan, J. Daunizeau, M.J. Rosa, K.J. Friston, T.M.Schofield, and A.P. Leff. Comparing families of dynamic causal models. *PLoS Comput Biol*, 6(3):e1000709, 2010. 145

Christophe Phillips, Michael D. Rugg, and Karl J. Friston. Anatomically informed basis functions for eeg source localization: Combining functional and anatomical constraints. *NeuroImage*, 16(3, Part 1):678 – 695, 2002. 9

D. Poeppel. A critical review of PET studies of phonological processing. *Brain Lang*, 55:317–351, Dec 1996. 3

J.B. Poline. Contrasts and classical inference. In R.S.J. Frackowiak, K.J. Friston, C. Frith, R. Dolan, K.J. Friston, C.J. Price, S. Zeki, J. Ashburner, and W.D. Penny, editors, *Human Brain Function*. Academic Press, 2nd edition, 2003. 16

J.B. Poline, A.P. Holmes, K.J. Worsley, and K.J. Friston. Making statistical inferences. In R.S.J. Frackowiak, K.J. Friston, C. Frith, R. Dolan, and J.C. Mazziotta, editors, *Human Brain Function*, pages 85–106. Academic Press USA, 1997. 16

M. I. Posner. Orienting of attention. *Q J Exp Psychol*, 32:3–25, Feb 1980. 83

M. Pyka, D. Heider, S. Hauke, T. Kircher, and A. Jansen. Dynamic causal modeling with genetic algorithms. *J. Neurosci. Methods*, 194:402–406, Jan 2011. 155, 162

A. Rauch, G. Rainer, and N. K. Logothetis. The effect of a serotonin-induced dissociation between spiking and perisynaptic activity on BOLD functional MRI. *Proc. Natl. Acad. Sci. U.S.A.*, 105:6759–6764, May 2008. 10, 28, 127

G. Rees, K. Friston, and C. Koch. A direct quantitative relationship between the functional properties of human and macaque V5. *Nat. Neurosci.*, 3:716–723, Jul 2000. 10, 27

J. J. Riera and A. Sumiyoshi. Brain oscillations: ideal scenery to understand the neurovascular coupling. *Curr. Opin. Neurol.*, 23:374–381, Aug 2010. 3, 10

J. J. Riera, X. Wan, J. C. Jimenez, and R. Kawashima. Nonlinear local electrovascular coupling. I: A theoretical model. *Hum Brain Mapp*, 27:896–914, Nov 2006. 101, 102, 103, 106, 107, 108, 114, 118, 128, 129, 130, 170, 171

J. J. Riera, J. C. Jimenez, X. Wan, R. Kawashima, and T. Ozaki. Nonlinear local electrovascular coupling. II: From data to neuronal masses. *Hum Brain Mapp*, 28: 335–354, Apr 2007. 20, 101, 115, 121

Jorge J. Riera, Arne Schousboe, Helle S. Waagepetersen, Clare Howarth, and Fahmeed Hyder. The micro-architecture of the cerebral cortex: Functional neuroimaging models and metabolism. *NeuroImage*, 40(4):1436 – 1459, 2008. 2

Petra Ritter and Arno Villringer. Simultaneous eeg-fmri. *Neuroscience and Biobehavioral Reviews*, 30(6):823 – 838, 2006. 7

B. P. Rogers, V. L. Morgan, A. T. Newton, and J. C. Gore. Assessing functional connectivity in the human brain by fMRI. *Magn Reson Imaging*, 25:1347–1357, Dec 2007. 22

M. J. Rosa, J. Daunizeau, and K. J. Friston. EEG-fMRI integration: a critical review of biophysical modeling and data analysis approaches. *J. Integr. Neurosci.*, 9:453–476, Dec 2010a. 7

M. J. Rosa, J. Kilner, F. Blankenburg, O. Josephs, and W. Penny. Estimating the transfer function from neuronal activity to BOLD using simultaneous EEG-fMRI. *Neuroimage*, 49:1496–1509, Jan 2010b. 8, 11, 101

M.J. Rosa, S. Bestmann, L. Harrison, and W. Penny. Bayesian Model Selection Maps for Group Studies. *NeuroImage*, 49(2):1496–1509, 2010c. 30

B. R. Rosen, R. L. Buckner, and A. M. Dale. Event-related functional MRI: past, present, and future. *Proc. Natl. Acad. Sci. U.S.A.*, 95:773–780, Feb 1998. 7

E. Rounis, K. E. Stephan, L. Lee, H. R. Siebner, A. Pesenti, K. J. Friston, J. C. Rothwell, and R. S. Frackowiak. Acute changes in frontoparietal activity after repetitive transcranial magnetic stimulation over the dorsolateral prefrontal cortex in a cued reaction time task. *J. Neurosci.*, 26:9629–9638, Sep 2006. 89

A. Salek-Haddadi, M. Merschhemke, L. Lemieux, and D.R. Fish. Simultaneous EEG-Correlated Ictal fMRI. *Neuroimage*, 16:32–40, May 2002. 49

M. A. Sato, T. Yoshioka, S. Kajihara, K. Toyama, N. Goda, K. Doya, and M. Kawato. Hierarchical Bayesian estimation for MEG inverse problem. *Neuroimage*, 23:806–826, Nov 2004. 40

T. M. Schofield, P. Iverson, S. J. Kiebel, K. E. Stephan, J. M. Kilner, K. J. Friston, J. T. Crinion, C. J. Price, and A. P. Leff. Changing meaning causes coupling changes within higher levels of the cortical hierarchy. *Proc. Natl. Acad. Sci. U.S.A.*, 106:11765–11770, Jul 2009. 26

G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):pp. 461–464, 1978. ISSN 00905364. 42

R. G. Shulman, F. Hyder, and D. L. Rothman. Cerebral energetics and the glycogen shunt: neurochemical basis of functional imaging. *Proc. Natl. Acad. Sci. U.S.A.*, 98:6417–6422, May 2001. 100

J. Sijbers, I. Michiels, M. Verhoye, J. Van Audekerke, A. Van der Linden, and D. Van Dyck. Restoration of mr-induced artifacts in simultaneously recorded mr/eeg data. *Magnetic Resonance Imaging*, 17(9):1383 – 1391, 1999. 7

M. Singh, S. Kim, and T.S. Kim. Correlation between BOLD-fMRI and EEG signal changes in response to visual stimulus frequency in humans. *Magn Reson Med*, 49: 108–114, Jan 2003. 68, 119

W. Skrandies. Visual information processing: topography of brain electrical activity. *Biological Psychology*, 40:1–15, 1995. 57

R. Sladky, K. J. Friston, J. Trostl, R. Cunnington, E. Moser, and C. Windischberger. Slice-timing effects and their correction in functional MRI. *Neuroimage*, 58:588–594, Sep 2011. 75

A. J. Smith, H. Blumenfeld, K. L. Behar, D. L. Rothman, R. G. Shulman, and F. Hyder. Cerebral energetics and spiking frequency: the neurophysiological basis of fMRI. *Proc. Natl. Acad. Sci. U.S.A.*, 99:10765–10770, Aug 2002. 10, 128

A. P. Smith, K. E. Stephan, M. D. Rugg, and R. J. Dolan. Task and content modulate amygdala-hippocampal connectivity in emotional retrieval. *Neuron*, 49:631–638, Feb 2006. 26

R. C. Sotero and N. J. Trujillo-Barreto. Modelling the role of excitatory and inhibitory neuronal activity in the generation of the BOLD signal. *Neuroimage*, 35:149–165, Mar 2007. 100

R. C. Sotero and N. J. Trujillo-Barreto. Biophysical model for integrating neuronal activity, EEG, fMRI and metabolism. *Neuroimage*, 39:290–309, Jan 2008. 78, 101, 161

R. C. Sotero, N. J. Trujillo-Barreto, Y. Iturria-Medina, F. Carbonell, and J. C. Jimenez. Realistically coupled neural mass models can generate EEG rhythms. *Neural Comput*, 19:478–512, Feb 2007. 20, 100

R. C. Sotero, N. J. Trujillo-Barreto, J. C. Jimenez, F. Carbonell, and R. Rodriguez-Rojas. Identification and comparison of stochastic metabolic/hemodynamic models (sMHM) for the generation of the BOLD signal. *J Comput Neurosci*, 26:251–269, Apr 2009. 101, 162

K. E. Stephan and W. D. Penny. Dynamic causal models and bayesian selection. In K. Friston, J. Ashburner, S. Kiebel, T. Nichols, and W. Penny, editors, *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier, London, 2007. 44, 81

K. E. Stephan, W. D. Penny, J. Daunizeau, R. J. Moran, and K. J. Friston. Bayesian model selection for group studies. *Neuroimage*, 46:1004–1017, Jul 2009. 40, 45, 81, 82, 87, 92, 93, 96, 129

K.E. Stephan and K.J. Friston. Analyzing effective connectivity with functional magnetic resonance imaging. *WIREs Cognitive Sience*, 1:446–459, 2010. doi: 10.1002/wcs.58. 22, 131, 135

K.E. Stephan, L. Harrison, S.J. Kiebel, O. David, W.D. Penny, and K.J. Friston. Dynamic causal models of neural system dynamics: current state and future extensions. *Journal of Biosciences*, pages 129–144, 2007a. 134

K.E. Stephan, N. Weiskopf, P.M. Drysdale, P.A. Robinson, and K.J. Friston. Comparing hemodynamic models with DCM. *NeuroImage*, 38:387–401, 2007b. 45

K.E. Stephan, L. Kasper, L. Harrison, J. Daunizeau, H. den Ouden, M. Breakspear, and K.J. Friston. Nonlinear dynamic causal models for FMRI. *NeuroImage*, 42(2): 649–662, 2008. 23, 133, 162

B. A. Strange, A. Duggins, W. Penny, R. J. Dolan, and K. J. Friston. Information theory, novelty and hippocampal responses: unpredicted or unpredictable? *Neural Netw*, 18:225–230, Apr 2005. 85

C. Summerfield and E. Koechlin. A neural representation of prior information during perceptual inference. *Neuron*, 59:336–347, Jul 2008. 26, 45

C. Summerfield, T. Egner, M. Greene, E. Koechlin, J. Mangels, and J. Hirsch. Predictive codes for forthcoming perception in the frontal cortex. *Science*, 314:1311–1314, Nov 2006. 26

J. Talairach and P. Tournoux. *Co-Planar Stereotaxic Atlas of the Human Brain*. Thieme Medical Publishers, 1988. 58, 84

C. Tallon-Baudry and O. Bertrand. Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, 3(4), 1999. 55

J. X. Tao, A. Ray, S. Hawes-Ebersole, and J. S. Ebersole. Intracranial EEG substrates of scalp EEG interictal spikes. *Epilepsia*, 46:669–676, May 2005. 5

K. Thomsen, N. Offenhauser, and M. Lauritzen. Principal neuron spiking: neither necessary nor sufficient for cerebral blood flow in rat cerebellum. *J. Physiol. (Lond.)*, 560:181–189, Oct 2004. 10, 27, 99

Keiichiro Toma, Takahiro Matsuoka, Ilka Immisch, Tatsuya Mima, Daniel Waldvogel, Benjamin Koshy, Takashi Hanakawa, Holly Shill, and Mark Hallett. Generators of movement-related cortical potentials: fmri-constrained eeg dipole source analysis. *NeuroImage*, 17(1):161 – 173, 2002. 9

Roberto Trotta. Applications of Bayesian model selection to cosmological parameters. *Monthly Notices of the Royal Astronomical Society*, 378(1):72–82, 2007. ISSN 1365-2966. doi: 10.1111/j.1365-2966.2007.11738.x. 132

N. Trujillo-Barreto, E. Martinez-Montes, L. Melie-Garcia, and P. Valdes-Sosa. A symmetrical Bayesian model for fMRI and EEG/MEG neuroimage fusion. *International Journal of Bioelectromagnetism*, 3, 2001. 50

N. J. Trujillo-Barreto, E. Aubert-Vazquez, and P. A. Valdes-Sosa. Bayesian model averaging in EEG/MEG imaging. *Neuroimage*, 21:1300–1319, Apr 2004. 35

R. Turner and T. Jones. Techniques for imaging neuroscience. *Br Med Bull*, 65:3–20, 2003. 2

P. A. Valdes-Sosa, J. M. Sanchez-Bornot, R. C. Sotero, Y. Iturria-Medina, Y. Aleman-Gomez, J. Bosch-Bayard, F. Carbonell, and T. Ozaki. Model driven EEG/fMRI fusion of brain oscillations. *Hum Brain Mapp*, 30:2701–2721, Sep 2009. 9, 101

S. Vanni, J. Warnking, M. Dojat, C. Delon-Martin, J. Bullier, and C. Segebarth. Sequence of pattern onset responses in the human visual areas: an fmri constrained vep source analysis. *NeuroImage*, 21(3):801 – 817, 2004. 8

R. Vigario, J. Särelä, V. Jousmäki, M. Hämäläinen, and E. Oja. Independent component approach to the analysis of EEG and MEG recordings. *IEEE Trans Biomed Eng*, 47:589–593, May 2000. 160

A. Viswanathan and R.D. Freeman. Neurometabolic coupling in cerebral cortex reflects synaptic more than spiking activity. *Nat. Neurosci.*, 10:1308–1312, Oct 2007. 10, 27, 99, 127

S. A. Teukolsky W. H. Press, B. P. Flannery and W. T. Vetterling, editors. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 2nd edition, 1992. 38

Eric-Jan Wagenmakers, Tom Lodewyckx, Himanshu Kuriyal, and Raoul Grasman. Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognitive Psychology*, 60(3):158 – 189, 2010. 132

M. Wagner and M. Fuchs. Integration of functional MRI, structural MRI, EEG and MEG. *Int Journal of Bioelectromag*, 2001. 7

X. Wan, J. Riera, K. Iwata, M. Takahashi, T. Wakabayashi, and R. Kawashima. The neural basis of the hemodynamic response nonlinearity in human primary visual

cortex: Implications for neurovascular coupling mechanism. *NeuroImage*, 32(3): 616–625, 2006. 50, 52, 53, 55, 68, 77, 119, 160

H. Wang, I. M. Hitron, C. Iadecola, and V. M. Pickel. Synaptic and vascular associations of neurons containing cyclooxygenase-2 and nitric oxide synthase in rat somatosensory cortex. *Cereb. Cortex*, 15:1250–1260, Aug 2005. 10, 100, 108, 129

Larry Wasserman. Bayesian model selection and model averaging. *Journal of Mathematical Psychology*, 44(1):92 – 107, 2000. 34

H. R. Wilsom. *Spikes, decisions, and actions: The dynamical foundations of neuroscience*. Oxford University Press, 1st edition, 1999. 20

Hugh R. Wilson and Jack D. Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, 12(1):1 – 24, 1972. 109

D. Wipf and S. Nagarajan. A unified Bayesian framework for MEG/EEG source imaging. *Neuroimage*, 44:947–966, Feb 2009. 42

M. W. Woolrich. Bayesian inference in fMRI. *Neuroimage*, 2011. submitted. 17

M. W. Woolrich, T. E. Behrens, and S. M. Smith. Constrained linear basis sets for HRF modelling using Variational Bayes. *Neuroimage*, 21:1748–1761, Apr 2004a. 40

M. W. Woolrich, M. Jenkinson, J. M. Brady, and S. M. Smith. Fully Bayesian spatio-temporal modeling of FMRI data. *IEEE Trans Med Imaging*, 23:213–231, Feb 2004b. 86

M. W. Woolrich, T. E. Behrens, C. F. Beckmann, and S. M. Smith. Mixture models with adaptive spatial regularization for segmentation with an application to FMRI data. *IEEE Trans Med Imaging*, 24:1–11, Jan 2005. 97

K. J. Worsley, A. C. Evans, S. Marrett, and P. Neelin. A three-dimensional statistical analysis for CBF activation studies in human brain. *J. Cereb. Blood Flow Metab.*, 12:900–918, Nov 1992. 14, 16

K. J. Worsley, S. Marrett, P. Neelin, A. C. Vandal, K. J. Friston, and A. C. Evans. A unified statistical approach for determining significant signals in images of cerebral activation. *Hum Brain Mapp*, 4:58–73, 1996. 14, 16

E. Yacoub, N. Harel, and K. Ugurbil. High-field fMRI unveils orientation columns in humans. *Proc. Natl. Acad. Sci. U.S.A.*, 105:10607–10612, Jul 2008. 2